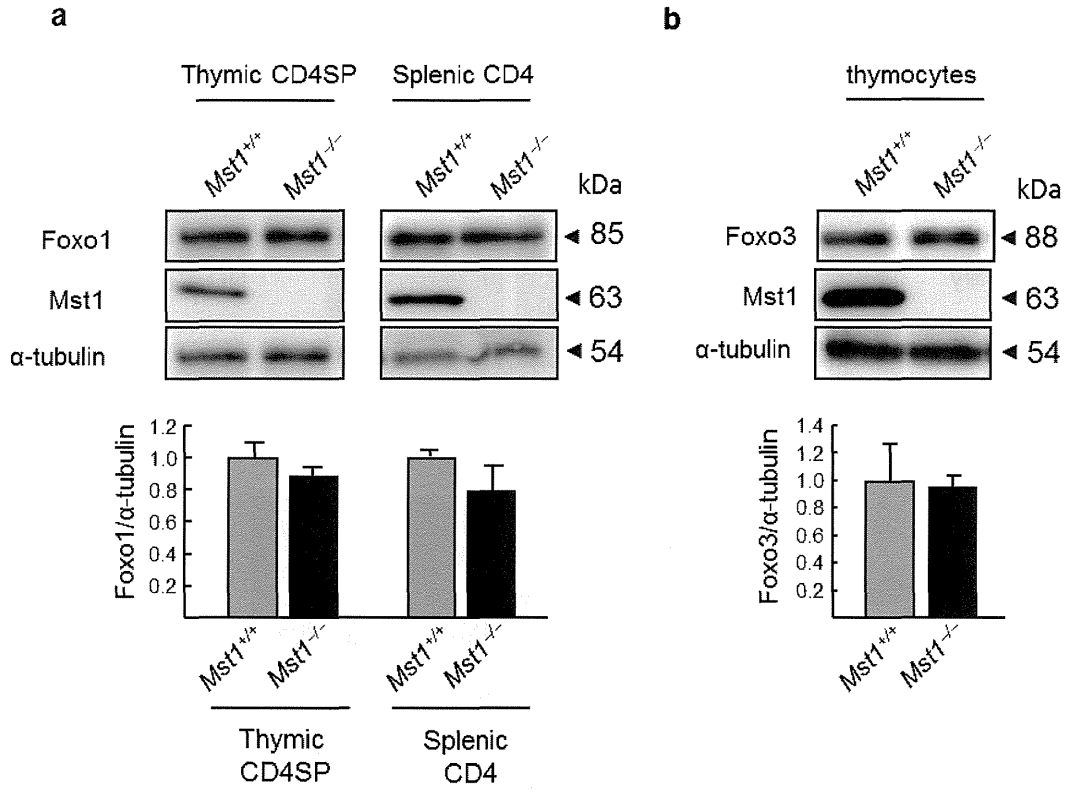
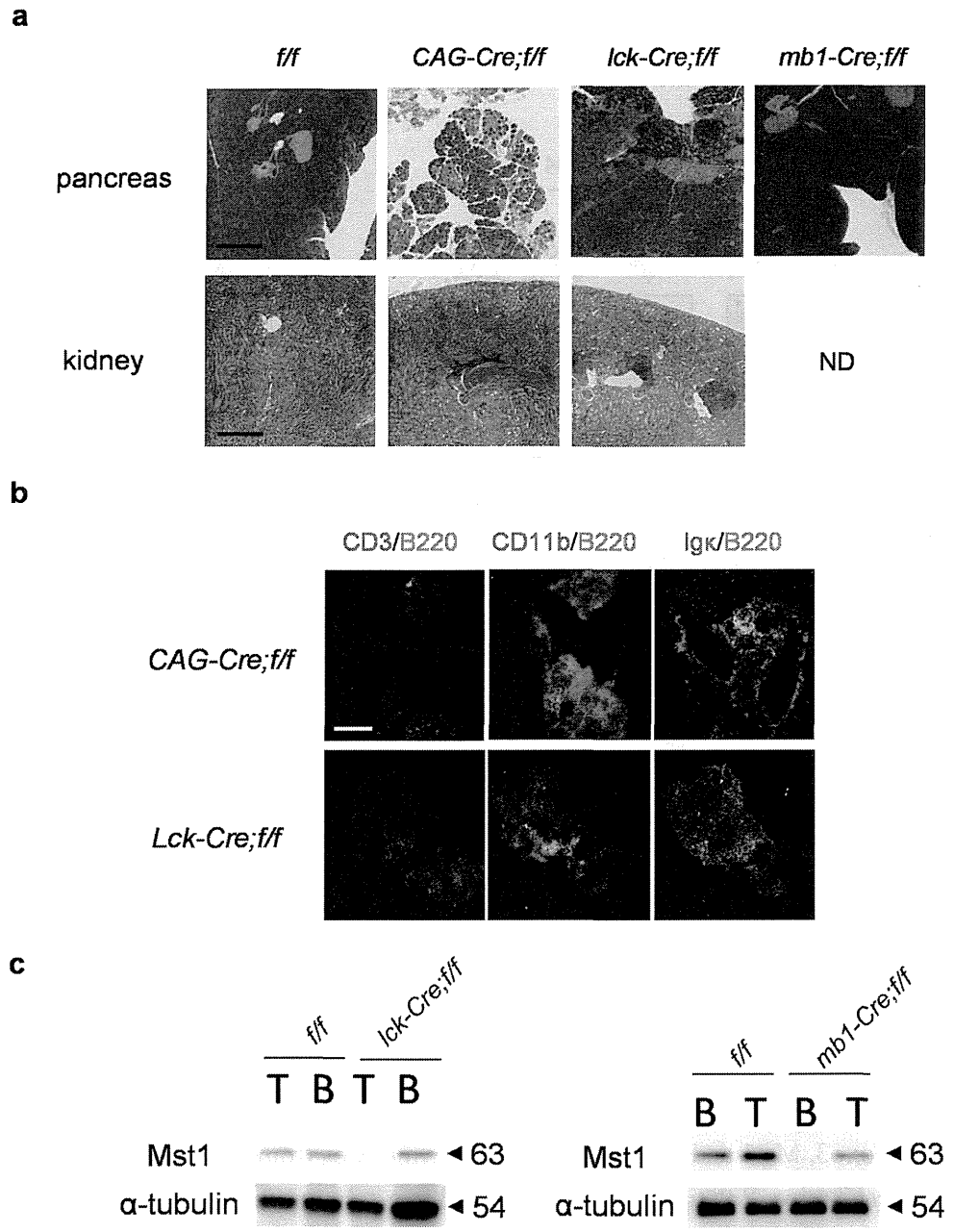


Supplementary Figure S5. Foxo1 and Foxo3 expression in thymocytes and splenic T cells. (**a**) Expression of Foxo1, Mst1 and α-tubulin in CD4SP thymocytes or splenic CD4 T cells from 2-3 month old *Mst1*$^{+/+}$ and *Mst1*$^{-/-}$ mice. The numbers indicate approximate molecular size of Foxo, Foxo3, Mst1, and α-tubulin. Relative intensities of Foxo1 to α-tubulin (the average ± range, n = 2) are also shown in the bar graphs (low panels). (**b**) Expression of Foxo3, Mst1 and α-tubulin in whole thymocytes from 2-3 month old *Mst1*$^{+/+}$ and *Mst1*$^{-/-}$ mice. Relative intensity of Foxo3 to α-tubulin (the average ± range, n = 2) are also shown in the bar graphs (low panels, n = 2).



Supplementary Figure S6. Leukocyte infiltration in one-year-old Mst1-deficient mice. (**a**) H&E staining of tissue sections of the kidney and pancreas from one-year-old *Mst1*$^{f/f}$, *CAG-Cre;Mst1*$^{f/f}$, *lck-Cre;Mst1*$^{f/f}$, *or mb-1-Cre;Mst1*$^{f/f}$ mice (scale bar 400 μm). (**b**) Cells infiltrating the liver were examined by immunostaining with anti-CD3, anti-B220, anti-CD11b and anti-Igκ antibodies (scale bar 200 μm). (**c**) T cell–specific and B cell–specific Mst1 deletions were confirmed by immunoblotting.

LETTER

doi:10.1038/nature12873

Genetics of rheumatoid arthritis contributes to biology and drug discovery

A list of authors and their affiliations appears at the end of the paper

A major challenge in human genetics is to devise a systematic strategy to integrate disease-associated variants with diverse genomic and biological data sets to provide insight into disease pathogenesis and guide drug discovery for complex traits such as rheumatoid arthritis (RA)[1]. Here we performed a genome-wide association study meta-analysis in a total of >100,000 subjects of European and Asian ancestries (29,880 RA cases and 73,758 controls), by evaluating ~10 million single-nucleotide polymorphisms. We discovered 42 novel RA risk loci at a genome-wide level of significance, bringing the total to 101 (refs 2–4). We devised an *in silico* pipeline using established bioinformatics methods based on functional annotation[5], *cis*-acting expression quantitative trait loci[6] and pathway analyses[7–9]—as well as novel methods based on genetic overlap with human primary immunodeficiency, haematological cancer somatic mutations and knockout mouse phenotypes—to identify 98 biological candidate genes at these 101 risk loci. We demonstrate that these genes are the targets of approved therapies for RA, and further suggest that drugs approved for other indications may be repurposed for the treatment of RA. Together, this comprehensive genetic study sheds light on fundamental genes, pathways and cell types that contribute to RA pathogenesis, and provides empirical evidence that the genetics of RA can provide important information for drug discovery.

We conducted a three-stage trans-ethnic meta-analysis (Extended Data Fig. 1). On the basis of the polygenic architecture of RA[10] and shared genetic risk among different ancestry[3,4], we proposed that combining a genome-wide association study (GWAS) of European and Asian ancestry would increase power to detect novel risk loci. In stage 1, we combined 22 GWAS for 19,234 cases and 61,565 controls of European and Asian ancestry[2–4]. We performed trans-ethnic, European-specific and Asian-specific GWAS meta-analysis by evaluating ~10 million single-nucleotide polymorphisms (SNPs)[11]. Characteristics of the cohorts, genotyping platforms and quality control criteria are described in Extended Data Table 1 (overall genomic control inflation factor $\lambda_{GC} < 1.075$).

Stage 1 meta-analysis identified 57 loci that satisfied a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$, including 17 novel loci (Extended Data Fig. 2). We then conducted a two-step replication study (stage 2 for *in silico* and stage 3 for *de novo*) in 10,646 RA cases and 12,193 controls for the loci with $P < 5.0 \times 10^{-6}$ in stage 1. In a combined analysis of stages 1–3, we identified 42 novel loci with $P < 5.0 \times 10^{-8}$ in any of the trans-ethnic, European or Asian meta-analyses. This increases the total number of RA risk loci to 101 (Table 1 and Supplementary Table 1).

Comparison of 101 RA risk loci revealed significant correlations of risk allele frequencies (RAFs) and odds ratios (ORs) between Europeans and Asians (Extended Data Fig. 3a–c; Spearman's $\rho = 0.67$ for RAF and 0.76 for OR; $P < 1.0 \times 10^{-13}$), although five loci demonstrated population-specific associations ($P < 5.0 \times 10^{-8}$ in one population but $P > 0.05$ in the other population without overlap of the 95% confidence intervals (95% CIs) of the ORs). In the population-specific genetic risk model, the 100 RA risk loci outside of the major histocompatibility complex (MHC) region[12] explained 5.5% and 4.7% of heritability in Europeans and Asians, respectively, with 1.6% of the heritability explained by the novel loci. The trans-ethnic genetic risk model, based on the RAF from one population but the OR from the other population, could explain the majority (>80%) of the known heritability in each population (4.7% for Europeans and 3.8% for Asians). These observations support our hypothesis that the genetic risk of RA is shared, in general, among Asians and Europeans.

We assessed enrichment of 100 non-MHC RA risk loci in epigenetic chromatin marks[13] (Extended Data Fig. 3d). Of 34 cell types investigated, we observed significant enrichment of RA risk alleles with trimethylation of histone H3 at lysine 4 (H3K4me3) peaks in primary $CD4^+$ regulatory T cells (T_{reg} cells; $P < 1.0 \times 10^{-5}$). For the RA risk loci enriched with T_{reg} H3K4me3 peaks, we incorporated the epigenetic annotations along with trans-ethnic differences in patterns of linkage disequilibrium to fine-map putative causal risk alleles (Extended Data Fig. 3e, f).

We found that approximately two-thirds of RA risk loci demonstrated pleiotropy with other human phenotypes (Extended Data Fig. 4), including immune-related diseases (for example, vitiligo, primary biliary cirrhosis), inflammation-related or haematological biomarkers (for example, fibrinogen, neutrophil counts) and other complex traits (for example, cardiovascular diseases).

Each of 100 non-MHC RA risk loci contains on average ~4 genes in the region of linkage disequilibrium (in total 377 genes). To prioritize systematically the most likely biological candidate gene, we devised an *in silico* bioinformatics pipeline. In addition to the published methods that integrate data across associated loci[7,8], we evaluated several biological data sets to test for enrichment of RA risk genes, which helps to pinpoint a specific gene in each loci (Extended Data Figs 5, 6 and Supplementary Tables 2–4).

We first conducted functional annotation of RA risk SNPs. Sixteen per cent of SNPs were in linkage disequilibrium with missense SNPs ($r^2 > 0.80$; Extended Data Fig. 5a, b). The proportion of missense RA risk SNPs was higher compared with a set of genome-wide common SNPs (8.0%), and relatively much higher in the explained heritability (~26.8%). Using *cis*-acting expression quantitative trait loci (*cis*-eQTL) data obtained from peripheral blood mononuclear cells (5,311 individuals)[6] and from $CD4^+$ T cells and $CD14^+CD16^-$ monocytes (212 individuals), we found that RA risk SNPs in 44 loci showed *cis*-eQTL effects (false discovery rate (FDR) q or permutation $P < 0.05$; Extended Data Table 2).

Second, we evaluated whether genes from RA risk loci overlapped with human primary immunodeficiency (PID) genes[14], and observed significant overlap (14/194 = 7.2%, $P = 1.2 \times 10^{-4}$; Fig. 1a and Extended Data Fig. 5c). Classification categories of PID genes showed different patterns of overlap: the highest proportion of overlap was in 'immune dysregulation' (4/21 = 19.0%, $P = 0.0033$) but there was no overlap in 'innate immunity'.

Third, we evaluated overlap with cancer somatic mutation genes[15], under the hypothesis that genes with cell growth advantages may contribute to RA development. Among 444 genes with registered cancer somatic mutations[15], we observed significant overlap with genes implicated in haematological cancers (17/251 = 6.8%, $P = 1.2 \times 10^{-4}$; Fig. 1b and Extended Data Fig. 5d), but not with genes implicated in non-haematological cancers (6/221 = 2.7%, $P = 0.56$).

Table 1 | Novel rheumatoid arthritis risk loci identified by trans-ethnic GWAS meta-analysis in >100,000 subjects

SNP	Chr	Genes	A1/A2 (+)	Trans-ethnic OR (95% CI)	Trans-ethnic P	European OR (95% CI)	European P	Asian OR (95% CI)	Asian P
rs227163	1	*TNFRSF9*	C/T	1.04 (1.02–1.06)	3.9×10^{-4}	1.00 (0.97–1.03)	9.3×10^{-1}	1.11 (1.08–1.16)*	3.1×10^{-9}*
rs28411352	1	*MTF1-INPP5B*	T/C	1.11 (1.08–1.14)*	2.8×10^{-12}*	1.10 (1.07–1.14)*	5.9×10^{-9}*	1.12 (1.06–1.19)	7.8×10^{-5}
rs2105325	1	*LOC100506023*	C/A	1.12 (1.08–1.15)*	6.9×10^{-13}*	1.12 (1.08–1.15)*	3.3×10^{-11}*	1.13 (1.04–1.23)	5.2×10^{-3}
rs10175798	2	*LBH*	A/G	1.08 (1.06–1.11)*	1.1×10^{-9}*	1.09 (1.06–1.12)*	4.2×10^{-8}*	1.07 (1.02–1.13)	6.4×10^{-3}
rs6732565	2	*ACOXL*	A/G	1.07 (1.05–1.10)*	2.7×10^{-8}*	1.10 (1.07–1.14)*	9.4×10^{-9}*	1.04 (1.00–1.08)	4.0×10^{-2}
rs6715284	2	*CFLAR-CASP8*	G/C	1.15 (1.10–1.20)*	1.8×10^{-9}*	1.15 (1.10–1.20)*	2.5×10^{-9}*	-	-
rs4452313	3	*PLCL2*	T/A	1.09 (1.06–1.12)*	1.6×10^{-10}*	1.11 (1.08–1.15)*	5.2×10^{-11}*	1.04 (0.99–1.09)	9.2×10^{-2}
rs3806624	3	*EOMES*	G/A	1.08 (1.05–1.11)*	8.6×10^{-9}*	1.08 (1.05–1.12)*	2.8×10^{-8}*	1.06 (0.99–1.14)	1.0×10^{-1}
rs9826828	3	*IL20RB*	A/G	1.44 (1.28–1.61)*	8.6×10^{-10}*	1.44 (1.28–1.61)*	8.7×10^{-10}*	-	-
rs13142500	4	*CLNK*	C/T	1.10 (1.07–1.13)*	3.0×10^{-9}*	1.10 (1.06–1.15)	2.4×10^{-6}	1.10 (1.04–1.15)	2.8×10^{-4}
rs2664035	4	*TEC*	A/G	1.07 (1.04–1.10)	9.5×10^{-8}	1.08 (1.05–1.11)*	3.3×10^{-8}*	1.03 (0.97–1.08)	3.3×10^{-1}
rs9378815	6	*IRF4*	C/G	1.09 (1.06–1.12)*	1.7×10^{-10}*	1.09 (1.05–1.12)	1.4×10^{-7}	1.10 (1.04–1.15)	2.3×10^{-4}
rs2234067	6	*ETV7*	C/A	1.15 (1.10–1.20)*	1.6×10^{-9}*	1.14 (1.09–1.19)*	4.1×10^{-8}*	1.22 (1.06–1.41)	7.0×10^{-3}
rs9373594	6	*PPIL4*	T/C	1.09 (1.06–1.12)*	3.0×10^{-9}*	1.07 (1.02–1.12)	6.5×10^{-3}	1.11 (1.07–1.15)*	4.8×10^{-8}*
rs67250450	7	*JAZF1*	T/C	1.10 (1.07–1.14)*	3.7×10^{-9}*	1.11 (1.07–1.14)*	2.6×10^{-9}*	1.02 (0.84–1.23)	8.5×10^{-1}
rs4272	7	*CDK6*	G/A	1.10 (1.06–1.13)*	5.0×10^{-9}*	1.10 (1.07–1.14)*	1.2×10^{-8}*	1.06 (0.98–1.15)	1.3×10^{-1}
rs998731	8	*TPD52*	T/C	1.08 (1.05–1.11)*	1.9×10^{-8}*	1.09 (1.06–1.12)*	6.6×10^{-9}*	1.02 (0.96–1.10)	4.9×10^{-1}
rs678347	8	*GRHL2*	G/A	1.08 (1.05–1.11)*	1.6×10^{-8}*	1.10 (1.06–1.13)*	7.3×10^{-9}*	1.03 (0.98–1.10)	2.6×10^{-1}
rs1516971	8	*PVT1*	T/C	1.15 (1.10–1.20)*	1.3×10^{-10}*	1.16 (1.11–1.21)*	3.2×10^{-11}*	-	-
rs12413578	10	10p14	C/T	1.20 (1.13–1.29)*	4.8×10^{-8}*	1.20 (1.12–1.29)	7.5×10^{-8}	-	-
rs793108	10	*ZNF438*	T/C	1.08 (1.05–1.10)*	1.3×10^{-9}*	1.07 (1.04–1.10)	6.1×10^{-7}	1.09 (1.04–1.14)	4.4×10^{-4}
rs2671692	10	*WDFY4*	A/G	1.07 (1.05–1.10)*	2.8×10^{-9}*	1.06 (1.03–1.09)	2.6×10^{-5}	1.10 (1.05–1.14)	9.9×10^{-6}
rs726288	10	*SFTPD*	T/C	1.14 (1.07–1.20)	1.6×10^{-5}	0.96 (0.86–1.06)	4.1×10^{-1}	1.22 (1.14–1.31)*	8.8×10^{-9}*
rs968567	11	*FADS1-FADS2-FADS3*	C/T	1.12 (1.07–1.16)*	1.8×10^{-8}*	1.12 (1.07–1.16)*	1.8×10^{-8}*	-	-
rs4409785	11	*CEP57*	C/T	1.12 (1.09–1.16)*	1.2×10^{-11}*	1.12 (1.08–1.16)*	3.6×10^{-9}*	1.16 (1.07–1.27)	4.3×10^{-4}
chr11:107967350	11	*ATM*	A/G	1.21 (1.13–1.29)*	1.4×10^{-8}*	1.21 (1.13–1.29)*	1.1×10^{-8}*	-	-
rs73013527	11	*ETS1*	C/T	1.09 (1.06–1.12)*	1.2×10^{-10}*	1.08 (1.05–1.11)	1.0×10^{-6}	1.14 (1.08–1.21)	4.1×10^{-6}
rs773125	12	*CDK2*	A/G	1.09 (1.06–1.12)*	1.1×10^{-10}*	1.09 (1.06–1.12)*	2.1×10^{-8}*	1.10 (1.04–1.17)	1.1×10^{-3}
rs10774624	12	*SH2B3-PTPN11*	G/A	1.09 (1.06–1.13)*	6.8×10^{-9}*	1.09 (1.06–1.13)*	6.9×10^{-9}*	-	-
rs9603616	13	*COG6*	C/T	1.10 (1.07–1.13)*	1.6×10^{-12}*	1.11 (1.07–1.14)*	2.8×10^{-11}*	1.08 (1.02–1.14)	1.0×10^{-2}
rs3783782	14	*PRKCH*	A/G	1.14 (1.09–1.18)*	2.2×10^{-9}*	1.12 (0.96–1.31)	1.4×10^{-1}	1.14 (1.09–1.19)*	4.4×10^{-9}*
rs1950897	14	*RAD51B*	T/C	1.10 (1.07–1.13)*	8.2×10^{-11}*	1.09 (1.06–1.12)*	5.0×10^{-8}*	1.16 (1.08–1.25)	1.1×10^{-4}
rs4780401	16	*TXNDC11*	T/G	1.07 (1.05–1.10)*	4.1×10^{-8}*	1.09 (1.06–1.13)*	8.7×10^{-9}*	1.03 (0.98–1.08)	2.5×10^{-1}
rs72634030	17	*C1QBP*	A/C	1.12 (1.08–1.17)*	1.5×10^{-9}*	1.12 (1.06–1.19)	2.9×10^{-5}	1.12 (1.07–1.18)	9.6×10^{-6}
rs1877030	17	*MED1*	C/T	1.09 (1.06–1.12)*	1.9×10^{-8}*	1.09 (1.05–1.13)	1.3×10^{-5}	1.09 (1.04–1.14)	3.2×10^{-4}
rs2469434	18	*CD226*	C/T	1.07 (1.05–1.10)*	8.9×10^{-10}*	1.05 (1.02–1.08)	6.7×10^{-4}	1.11 (1.07–1.15)*	1.2×10^{-8}*
chr19:10771941	19	*ILF3*	C/T	1.47 (1.30–1.67)*	8.6×10^{-10}*	1.47 (1.30–1.67)*	8.8×10^{-10}*	-	-
rs73194058	21	*IFNGR2*	C/A	1.08 (1.05–1.12)	1.2×10^{-6}	1.13 (1.08–1.18)*	2.6×10^{-8}*	1.03 (0.98–1.08)	2.9×10^{-1}
rs1893592	21	*UBASH3A*	A/C	1.11 (1.08–1.14)*	7.2×10^{-12}*	1.11 (1.07–1.15)*	9.8×10^{-9}*	1.11 (1.05–1.18)	1.3×10^{-4}
rs11089637	22	*UBE2L3-YDJC*	C/T	1.08 (1.05–1.11)*	2.1×10^{-9}*	1.10 (1.06–1.15)	2.0×10^{-7}	1.06 (1.02–1.10)	8.9×10^{-4}
rs909685	22	*SYNGR1*	A/T	1.13 (1.10–1.16)*	1.4×10^{-16}*	1.11 (1.08–1.15)*	6.4×10^{-12}*	1.23 (1.14–1.33)	2.0×10^{-7}
chrX:78464616	X	*P2RY10*	A/C	1.11 (1.07–1.15)*	3.5×10^{-8}*	1.16 (0.78–1.75)	4.6×10^{-1}	1.11 (1.07–1.15)*	3.6×10^{-8}*

SNPs newly associated with $P < 5.0 \times 10^{-8}$ in the combined study of the stage 1 GWAS meta-analysis and the stages 2 and 3 replication studies of trans-ethnic (Europeans and Asians), European or Asian ancestry are indicated. SNPs, positions and alleles are based on the positive (+) strand of NCBI build 37. A1 represents an RA risk allele. Chr, chromosome; OR, odds ratio; 95% CI, 95% confidence interval. Full results of the studies are available in Supplementary Table 1. Hyphens between gene names indicate that several candidate RA risk genes were included in the region.
*Association results with $P < 5.0 \times 10^{-8}$.

Fourth, we evaluated overlap with genes implicated in knockout mouse phenotypes[16]. Among the 30 categories of phenotypes[16], we observed 3 categories significantly enriched with RA risk genes ($P < 0.05/30 = 0.0017$): 'haematopoietic system phenotype', 'immune system phenotype', and 'cellular phenotype' (Extended Data Fig. 5e).

Last, we conducted molecular pathway enrichment analysis (Fig. 1c and Extended Data Fig. 5f). We observed enrichment (FDR $q < 0.05$) for T-cell-related pathways, consistent with cell-specific epigenetic marks, as well as enrichment for B-cell and cytokine signalling pathways (for example, interleukin (IL)-10, interferon, granulocyte–macrophage colony-stimulating factor (GM-CSF)). For comparison, our previous RA GWAS meta-analysis[2] did not identify the B-cell and cytokine signalling pathways, thereby indicating that as more loci are discovered, further biological pathways are identified.

On the basis of these new findings, we adopted the following 8 criteria to prioritize each of the 377 genes from the 100 non-MHC RA risk loci (Fig. 2 and Extended Data Fig. 6a–c): (1) genes with RA risk missense variant ($n = 19$); (2) *cis*-eQTL genes ($n = 51$); (3) genes prioritized by PubMed text mining[7] ($n = 90$); (4) genes prioritized by protein–protein interaction (PPI)[8] ($n = 63$); (5) PID genes ($n = 15$); (6) haematological cancer somatic mutation genes ($n = 17$); (7) genes prioritized by associated knockout mouse phenotypes ($n = 86$); and (8) genes prioritized by molecular pathway analysis[9] ($n = 35$).

Ninety-eight genes (26.0%) had a score ≥2, which we defined as 'candidate biological RA risk genes'. Nineteen loci included multiple biological RA risk genes (for example, *IL3* and *CSF2* at chromosome 5q31), whereas no biological gene was selected from 40 loci (Supplementary Table 5).

To provide empirical evidence of the pipeline, we evaluated relationships of the gene scores to independent genomic or epigenetic information. Genes with higher biological scores were more likely to be the nearest gene to the risk SNP (18.6% for gene score <2 and 49.0% for gene score ≥2; $P = 2.1 \times 10^{-8}$), and also to be included in the region where RA risk SNPs were overlapping with H3K4me3 T_{reg} peaks (41.9% for gene score <2 and 57.1% for gene score ≥2; $P = 0.034$). Further, T_{reg} cells demonstrated the largest increase in overlapping proportions with H3K4me3 peaks for increase of biological gene scores compared with other cell types (Extended Data Fig. 6d).

Finally, we evaluated the potential role of RA genetics in drug discovery. We proposed that if human genetics is useful for drug target validation, then it should identify existing approved drugs for RA. To test this 'therapeutic hypothesis'[1], we obtained 871 drug target genes corresponding to approved, in clinical trials or experimental drugs for human diseases[17,18] (Supplementary Table 6). We evaluated whether any of the protein products from the identified biological RA risk genes, or any genes from a direct PPI network with such protein products

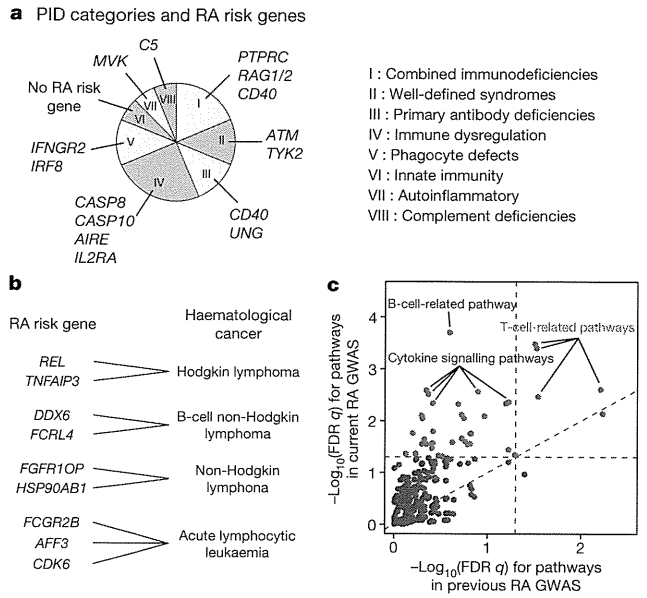


Figure 1 | Overlap of RA risk loci with PID genes, haematological cancer somatic mutations and molecular pathways. a, Overlap of RA risk genes with PID genes, subdivided by PID categories (I–VIII). **b**, Examples of overlap of haematological cancer somatic mutation genes with RA risk genes. **c**, Comparisons of molecular pathway analysis results between the current trans-ethnic meta-analysis (*y*-axis) and the previous meta-analysis for RA (*x*-axis)[2]. Each dot represents a molecular pathway. Dotted line represents FDR $q = 0.05$ or $y = x$.

(Fig. 3a), are the pharmacologically active targets of approved RA drugs (Extended Data Fig. 7a).

Twenty-seven drug target genes of approved RA drugs demonstrated significant overlap with 98 biological RA risk genes and 2,332 genes from the expanded PPI network (18 genes overlapped; 3.7-fold enrichment by permutation analysis, $P < 1.0 \times 10^{-5}$; Fig. 3b). For comparison, all drug target genes (regardless of disease indication) overlapped with 247 genes, which is 1.7-fold more enrichment than expected by chance, but less than 2.2-fold enrichment compared with overlap of the target genes of RA drugs ($P = 0.0035$). Examples of approved RA therapies identified by this analysis include tocilizumab[19,20] (anti-IL6R), tofacitinib[21] (JAK3 inhibitor) and abatacept[21] (CTLA4–immunoglobulin; Fig. 3c and Extended Data Fig. 8).

We also assessed how approved drugs for other diseases might be connected to biological RA risk genes. We highlight *CDK6* and *CDK4*, targets of three approved drugs for different types of cancer[22] (Fig. 3d).

RA risk SNP (cytoband)	Gene	Score	RA risk missense variant	*cis*-eQTL	PubMed text mining	PPI	PID	Haematological cancer	Knockout mouse phenotype	Molecular pathway	Nearest gene from RA risk SNP	T_{reg} primary cells	CD4+ memory primary cells	CD4+ naive primary cells	CD8+ memory primary cells	CD8+ naive primary cells	CD34+ primary cells	CD34+ cultured cells	Mobilized CD34+ primary cells	CD19+ primary cells	CD3+ primary cells	Drug target gene	RA drug target gene	PPI with RA drug target gene
			Biological gene criteria									Overlap with H3K4me3 peaks												
chr1:2523811 (1p36)	*TNFRSF14*	4																						
rs2301888 (1p36)	*PADI4*	2																						
rs2476601 (1p13)	*PTPN22*	5																						
rs2228145 (1q21)	*IL6R*	5																						
chr1:161644258 (1q23)	*FCGR2B*	5																						
rs17668708 (1q31)	*PTPRC*	6																						
rs34695944 (2p16-p15)	*REL*	4																						
rs9653442 (2q11)	*AFF3*	4																						
rs11889341 (2q32)	*STAT4*	3																						
rs6715284 (2q33)	*CFLAR*	3																						
rs1980422 (2q33)	*CD28*	4																						
rs3087243 (2q33)	*CTLA4*	4																						
rs45475795 (4q26-q27)	*IL2*	5																						
rs657075 (5q31)	*IL3*	4																						
rs657075 (5q31)	*CSF2*	4																						
rs2233424 (6p21)	*NFKBIE*	4																						
rs7752903 (6q23)	*TNFAIP3*	6																						
rs1571878 (6q27)	*CCR6*	2																						
rs4272 (7q21)	*CDK6*	4																						
chr7:128580042 (7q32)	*IRF5*	4																						
rs10985070 (9q33)	*TRAF1*	4																						
rs10985070 (9q33)	*C5*	4																						
rs706778 (10p15)	*IL2RA*	5																						
rs331463 (11p12)	*TRAF6*	4																						
rs331463 (11p12)	*RAG1*	4																						
rs508970 (11q12)	*CD5*	4																						
chr11:107967350 (11q22)	*ATM*	4																						
rs773125 (12q13)	*CDK2*	3																						
rs1633360 (12q13-q14)	*CDK4*	3																						
rs10774624 (12q24)	*SH2B3*	5																						
chr17:38031857 (17q12-q21)	*IKZF3*	4																						
chr17:38031857 (17q12-q21)	*CSF3*	4																						
rs8083786 (18p11)	*PTPN2*	3																						
rs34536443 (19p13)	*ICAM1*	4																						
rs34536443 (19p13)	*TYK2*	6																						
rs4239702 (20q13)	*CD40*	6																						
rs73194058 (21q22)	*IFNGR2*	6																						
rs2236668 (21q22)	*ICOSLG*	5																						
rs2236668 (21q22)	*AIRE*	4																						
rs3218251 (22q12)	*IL2RB*	3																						
rs5987194 (Xq28)	*IRAK1*	3																						

Figure 2 | Prioritized biological RA risk genes. Representative biological RA risk genes. We list the summary gene score derived from individual criteria (filled red box indicates criterion satisfied; 98 genes with a score ≥2 out of 377 genes included in the RA risk loci were defined as 'biological candidate genes'; see Extended Data Fig. 6). Filled blue boxes indicate the nearest gene to the RA risk SNP. Filled green boxes indicate overlap with H3K4me3 peaks in immune-related cells. Filled purple boxes indicate overlap with drug target genes. For full results, see Supplementary Table 5.

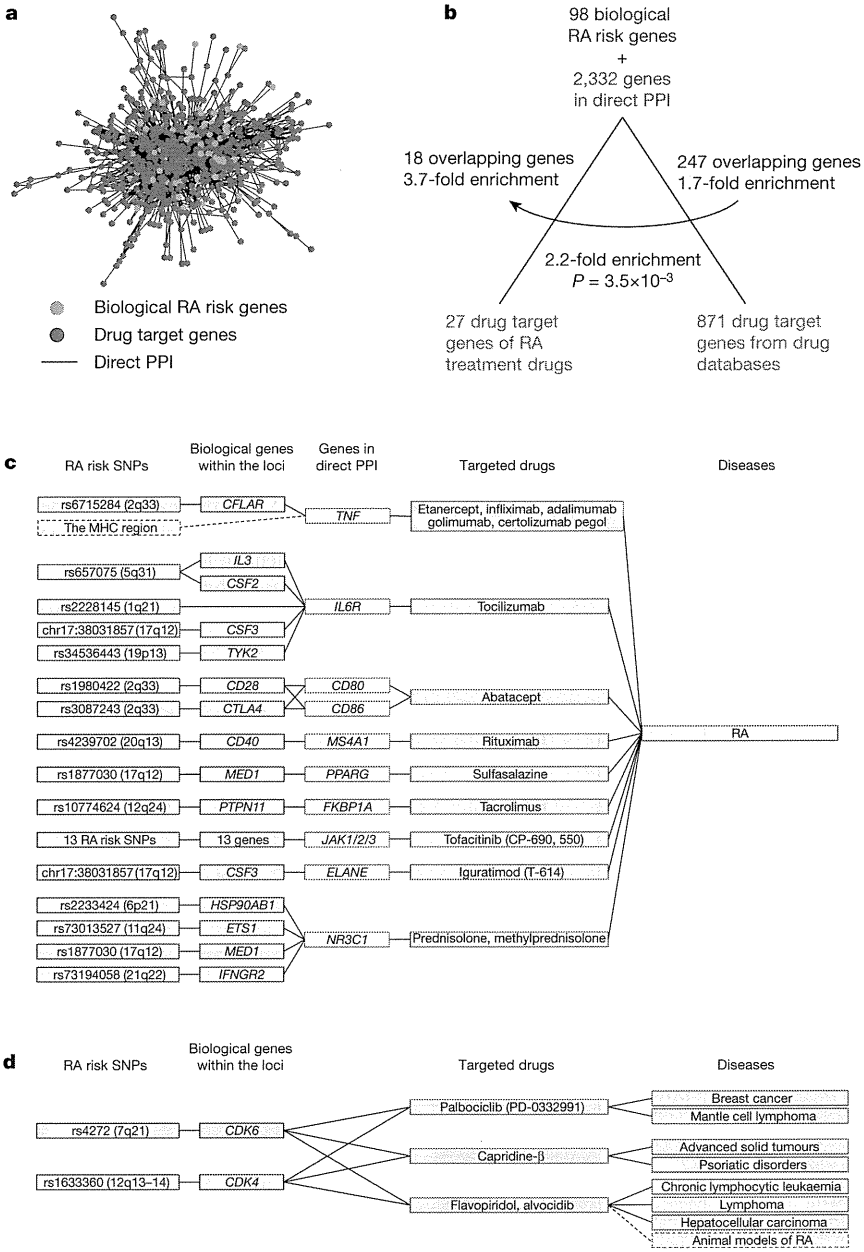


Figure 3 | Connection of biological RA risk genes to drug targets. a, PPI network of biological RA risk genes and drug target genes. **b**, Overlap and relative enrichment of 98 biological RA risk genes with targets of approved RA drugs and with all drug target genes. Enrichment was more apparent than that from all 377 RA risk genes (Extended Data Fig. 7c). **c**, Connections between RA risk SNPs (blue), biological genes (purple), genes from PPI (green) and approved RA drugs (orange). For full results, see Extended Data Fig. 8. **d**, Connections between RA genes and drugs indicated for other diseases.

In support for repurposing, one *CDK6/CDK4* inhibitor, flavopiridol, has been shown to ameliorate disease activity in animal models of RA[22]. Further, the biology is plausible, as several approved RA drugs were initially developed for cancer treatment and then repurposed for RA (for example, rituximab). Although further investigations are necessary, we propose that target genes/drugs selected by this approach could represent promising candidates for novel drug discovery for RA treatment.

We note that a non-random distribution of drug-to-disease indications in the databases could potentially bias our results. Namely, because RA risk genes are enriched for genes with immune function, spurious enrichment with drug targets could occur if the majority of drug indications in databases were for immune-mediated diseases or immune-related target genes. However, such enrichment was not evident in our analysis (~11% for drug indications and ~9% for target genes; Extended Data Fig. 7b).

Through a comprehensive genetic study with >100,000 subjects, we identified 42 novel RA risk loci and provided novel insight into RA pathogenesis. We particularly highlight the role of genetics for drug discovery. Although there have been anecdotal examples of this[1,23], our study provides a systematic approach by which human genetic data can be efficiently integrated with other biological information to derive biological insights and drive drug discovery.

METHODS SUMMARY

Details can be found in Methods, Extended Data Fig. 1, Extended Data Table 1 and Supplementary Information, including (1) information about the patient collections;

(2) genotyping, quality control and genotype imputation of GWAS data; (3) genome-wide meta-analysis (stage 1); (4) *in silico* and *de novo* replication studies (stages 2 and 3); (5) trans-ethnic and functional annotations of RA risk SNPs; (6) prioritization of biological candidate genes; and (7) drug target gene enrichment analysis.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 June; accepted 7 November 2013.
Published online 25 December 2013.

1. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nature Rev. Drug Discov.* **12,** 581–594 (2013).
2. Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet.* **42,** 508–514 (2010).
3. Okada, Y. *et al.* Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nature Genet.* **44,** 511–516 (2012).
4. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genet.* **44,** 1336–1340 (2012).
5. Ferreira, R. C. *et al.* Functional *IL6R* 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet.* **9,** e1003444 (2013).
6. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genet.* **45,** 1238–1243 (2013).
7. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5,** e1000534 (2009).
8. Rossin, E. J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7,** e1001273 (2011).
9. Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).
10. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44,** 483–489 (2012).
11. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
12. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genet.* **44,** 291–296 (2012).
13. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genet.* **45,** 124–130 (2013).
14. Parvaneh, N., Casanova, J. L., Notarangelo, L. D. & Conley, M. E. Primary immunodeficiencies: a rapidly evolving story. *J. Allergy Clin. Immunol.* **131,** 314–323 (2013).
15. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950 (2011).
16. Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A. & Richardson, J. E. The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* **40,** D881–D886 (2012).
17. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39,** D1035–D1041 (2011).
18. Zhu, F. *et al.* Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* **40,** D1128–D1136 (2012).
19. Smolen, J. S. *et al.* Consensus statement on blocking the effects of interleukin-6 and in particular by interleukin-6 receptor inhibition in rheumatoid arthritis and other inflammatory conditions. *Ann. Rheum. Dis.* **72,** 482–492 (2013).
20. Nishimoto, N. *et al.* Study of active controlled tocilizumab monotherapy for rheumatoid arthritis patients with an inadequate response to methotrexate (SATORI): significant reduction in disease activity and serum vascular endothelial growth factor by IL-6 receptor inhibition therapy. *Mod. Rheumatol.* **19,** 12–19 (2009).
21. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* **365,** 2205–2219 (2011).
22. Sekine, C. *et al.* Successful treatment of animal models of rheumatoid arthritis with small-molecule cyclin-dependent kinase inhibitors. *J. Immunol.* **180,** 1954–1961 (2008).
23. Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nature Biotechnol.* **30,** 317–320 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements R.M.P. is supported by National Institutes of Health (NIH) grants R01-AR057108, R01-AR056768, U01-GM092691 and R01-AR059648, and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. Y.O. is supported by a grant from the Japan Society of the Promotion of Science. D.W. is supported by a grant from the Australian National Health and Medical Research Council (1036541). G.T. is supported by the Rubicon grant from the Netherlands Organization for Scientific Research. A.Z. is supported by a grant from the Dutch Reumafonds (11-1-101) and from the Rosalind Franklin Fellowship, University of Groningen. S.-C.B., S.-Y.B. and H.-S.L. are supported by the Korea Healthcare technology R&D project, Ministry for Health and Welfare (A121983). J.M., M.A.G.-G. and L.R.-R. are funded by the RETICS program, RIER, RD12/0009 from the Instituto de Salud Carlos III, Health Ministry. S.R.-D. and L.Ä.'s work is supported by the Medical Biobank of Northern Sweden. H.K.C. is supported by NIH (NIAMS) grants R01-AR056291, R01-AR065944, R01-AR056768, P60 AR047785 and R21 AR056042. L.P. and L.K. are supported by a senior investigator grant from the European Research Council. S.R. is supported by NIH grants R01AR063759-01A1 and K08-KAR055688A. P.M.V. is a National Health and Medical Research Council Senior Principal Research Fellow. M.A.B. is funded by the National Health and Medical Research Foundation Senior Principal Research Fellowship, and a Queensland State Government Premier's Fellowship. H.X. is funded by the China Ministry of Science and Technology (973 program grant 2011CB946100), the National Natural Science Foundation of China (grants 30972339, 81020108029 and 81273283), and the Science and Technology Commission of Shanghai Municipality (grants 08XD1400400, 11410701600 and 10JC1418400). K.A.S. is supported by a Canada Research Chair, The Sherman Family Chair in Genomics Medicine, Canadian Institutes for Health Research grant 79321 and Ontario Research Fund grant 05-075. S.M. is supported by Health and Labour Sciences Research Grants. The BioBank Japan Project is supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government. This study is supported by the BE THE CURE (BTCure) project. We thank K. Akari, K. Tokunaga and N. Nishida for supporting the study.

Author Contributions Y.O. carried out the primary data analyses. D.W. managed drug target gene data. G.T. conducted histone mark analysis. T.R., H.-J.W., T.E., A.M., B.E.S., P.L.D. and L.F. conducted eQTL analysis. C.T., K.I., Y.K., K.O., A.S., S.Y., G.X., E.K. and K.A.S. conducted the *de novo* replication study. R.R.G., A.M., W.O., T.B., T.W.B., L.J., J. Yin, L.Y., D.-F.S., J. Yang, P.M.V., M.A.B. and H.X. conducted the *in silico* replication study. E.A.S., D.D., J.C., T.K., R.Y. and A.T. managed GWAS data. All other authors, as well as the members of the RACI and GARNET consortia, contributed to additional analyses and genotype and clinical data enrolments. Y.O. and R.M.P. designed the study and wrote the manuscript, with contributions from all authors on the final version of the manuscript.

Author Information Summary statistics from the GWAS meta-analysis, source codes, and data sources used in this study are available at http://plaza.umin.ac.jp/~yokada/datasource/software.htm. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.P. (robert.plenge@merck.com) or Y.O. (yokada.brc@tmd.ac.jp).

Yukinori Okada[1,2,3], Di Wu[1,2,3,4,5], Gosia Trynka[1,2,3], Towfique Raj[2,3,6], Chikashi Terao[7,8], Katsunori Ikari[9], Yuta Kochi[10], Koichiro Ohmura[8], Akari Suzuki[10], Shinji Yoshida[9], Robert R. Graham[11], Arun Manoharan[11], Ward Ortmann[11], Tushar Bhangale[11], Joshua C. Denny[12,13], Robert J. Carroll[12], Anne E. Eyler[13], Jeffrey D. Greenberg[14], Joel M. Kremer[15], Dimitrios A. Pappas[16], Lei Jiang[17], Jian Yin[17], Lingying Ye[17], Ding-Feng Su[18], Jian Yang[19,20], Gang Xie[21,22,23], Ed Keystone[24], Harm-Jan Westra[25], Tõnu Esko[3,26,27], Andres Metspalu[26], Xuezhong Zhou[28], Namrata Gupta[3], Daniel Mirel[3], Eli A. Stahl[29], Dorothée Diogo[1,2,3], Jing Cui[1,2,3], Katherine Liao[1,2,3], Michael H. Guo[1,3,27], Keiko Myouzen[10], Takahisa Kawaguchi[7], Marieke J. H. Coenen[30], Piet L. C. M. van Riel[31], Mart A. F. J. van de Laar[32], Henk-Jan Guchelaar[33], Tom W. J. Huizinga[34], Philippe Dieudé[35,36], Xavier Mariette[37], S. Louis Bridges Jr[38], Alexandra Zhernakova[25,34], Rene E. M. Toes[34], Paul P. Tak[39,40,41], Corinne Miceli-Richard[37], So-Young Bang[42], Hye-Soon Lee[42], Javier Martin[43], Miguel A. Gonzalez-Gay[44], Luis Rodriguez-Rodriguez[45], Solbritt Rantapää-Dahlqvist[46,47], Lisbeth Ärlestig[46,47], Hyon K. Choi[48,49,50], Yoichiro Kamatani[51], Pilar Galan[52], Mark Lathrop[53], the RACI consortium†, the GARNET consortium†, Steve Eyre[54,55], John Bowes[54,55], Anne Barton[54], Niek de Vries[56], Larry W. Moreland[57], Lindsey A. Criswell[58], Elizabeth W. Karlson[1], Atsuo Taniguchi[9], Ryo Yamada[59], Michiaki Kubo[60], Jun S. Liu[4], Sang-Cheol Bae[42], Jane Worthington[54,55], Leonid Padyukov[61], Lars Klareskog[61], Peter K. Gregersen[62], Soumya Raychaudhuri[1,2,3,63], Barbara E. Stranger[64,65], Philip L. De Jager[2,3,6], Lude Franke[25], Peter M. Visscher[19,20], Matthew A. Brown[19], Hisashi Yamanaka[9], Tsuneyo Mimori[8], Atsushi Takahashi[66], Huji Xu[17], Timothy W. Behrens[11], Katherine A. Siminovitch[21,22,23], Shigeki Momohara[9], Fumihiko Matsuda[7,67,68], Kazuhiko Yamamoto[10,69] & Robert M. Plenge[1,2,3]

[1]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [4]Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, USA. [5]Centre for Cancer Research, Monash Institute of Medical Research, Monash University, Clayton, Victoria 3800, Australia. [6]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [7]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. [8]Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan. [9]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo 162-0054, Japan. [10]Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan. [11]Immunology Biomarkers Group, Genentech, South San Francisco, California 94080, USA. [12]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. [13]Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. [14]New York University Hospital for Joint Diseases, New York, New York 10003, USA. [15]Department of Medicine, Albany Medical Center and The Center for Rheumatology, Albany, New York 12206, USA.

[16]Division of Rheumatology, Department of Medicine, New York, Presbyterian Hospital, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. [17]Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, Second Military Medical University, Shanghai 200003, China. [18]Department of Pharmacology, Second Military Medical University, Shanghai 200433, China. [19]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4072, Australia. [20]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia. [21]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [22]Toronto General Research Institute, Toronto, Ontario M5G 2M9, Canada. [23]Department of Medicine, University of Toronto, Toronto, Ontario M5S 2J7, Canada. [24]Department of Medicine, Mount Sinai Hospital and University of Toronto, Toronto M5S 2J7, Canada. [25]Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, Groningen 9700 RB, the Netherlands. [26]Estonian Genome Center, University of Tartu, Riia 23b, Tartu 51010, Estonia. [27]Division of Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA. [28]School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China. [29]The Department of Psychiatry at Mount Sinai School of Medicine, New York, New York 10029, USA. [30]Department of Human Genetics, Radboud University Medical Centre, Nijmegen 6500 HB, the Netherlands. [31]Department of Rheumatology, Radboud University Medical Centre, Nijmegen 6500 HB, the Netherlands. [32]Department of Rheumatology and Clinical Immunology, Arthritis Center Twente, University Twente & Medisch Spectrum Twente, Enschede 7500 AE, the Netherlands. [33]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands. [34]Department of Rheumatology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands. [35]Service de Rhumatologie et INSERM U699 Hôpital Bichat Claude Bernard, Assistance Publique des Hôpitaux de Paris, Paris 75018, France. [36]Université Paris 7-Diderot, Paris 75013, France. [37]Institut National de la Santé et de la Recherche Médicale (INSERM) U1012, Université Paris-Sud, Rhumatologie, Hôpitaux Universitaires Paris-Sud, Assistance Publique-Hôpitaux de Paris (AP-HP), Le Kremlin Bicêtre 94275, France. [38]Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [39]AMC/University of Amsterdam, Amsterdam 1105 AZ, the Netherlands. [40]GlaxoSmithKline, Stevenage SG1 2NY, UK. [41]University of Cambridge, Cambridge CB2 1TN, UK. [42]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul 133-792, South Korea. [43]Instituto de Parasitologia y Biomedicina Lopez-Neyra, CSIC, Granada 18100, Spain. [44]Department of Rheumatology, Hospital Marques de Valdecilla, IFIMAV, Santander 39008, Spain. [45]Hospital Clinico San Carlos, Madrid 28040, Spain. [46]Department of Public Health and Clinical Medicine, Umeå University, Umeå SE-901 87, Sweden. [47]Department of Rheumatology, Umeå University, Umeå SE-901 87, Sweden. [48]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston 02115, Massachusetts, USA. [49]Section of Rheumatology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [50]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [51]Centre d'Etude du Polymorphisme Humain (CEPH), Paris 75010, France. [52]Université Paris 13 Sorbonne Paris Cité, UREN (Nutritional Epidemiology Research Unit), Inserm (U557), Inra (U1125), Cnam, Bobigny 93017, France. [53]McGill University and Génome Québec Innovation Centre, Montréal, Québec H3A 0G1 Canada. [54]Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester M13 9NT, UK. [55]National Institute for Health Research, Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals National Health Service Foundation Trust, Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK. [56]Department of Clinical Immunology and Rheumatology & Department of Genome Analysis, Academic Medical Center/University of Amsterdam, Amsterdam 1105 AZ, the Netherlands. [57]Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. [58]Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, Department of Medicine, University of California San Francisco, San Francisco, California 94117, USA. [59]Unit of Statistical Genetics, Center for Genomic Medicine Graduate School of Medicine Kyoto University, Kyoto 606-8507, Japan. [60]Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan. [61]Rheumatology Unit, Department of Medicine (Solna), Karolinska Institutet, Stockholm SE-171 76, Sweden. [62]The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York 11030, USA. [63]NIHR Manchester Musculoskeletal Biomedical, Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK. [64]Section of Genetic Medicine, University of Chicago, Chicago, Illinois 60637, USA. [65]Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. [66]Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan. [67]Core Research for Evolutional Science and Technology (CREST) program, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. [68]Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. [69]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-0033, Japan. †Lists of participants and their affiliations appear in the Supplementary Information.

METHODS

Subjects. Our study included 29,880 RA cases (88.1% seropositive and 9.3% seronegative for anti-citrullinated peptide antibody (ACPA) or rheumatoid factor (RF), and 2.6% who had unknown autoantibody status) and 73,758 controls. All RA cases fulfilled the 1987 criteria of the American College of Rheumatology for RA diagnosis[24], or were diagnosed with RA by a professional rheumatologist. The 19,234 RA cases and 61,565 controls enrolled in the stage 1 trans-ethnic GWAS meta-analysis were obtained from 22 studies on people with European and Asian ancestries (14,361 RA cases and 43,923 controls from 18 studies of Europeans and 4,873 RA cases and 17,642 controls from 4 studies of Asians): BRASS[2], CANADA[2], EIRA[2], NARAC1[2], NARAC2[2], WTCCC[2], Rheumatoid Arthritis Consortium International for Immunochip (RACI)-UK[4], RACI-US[4], RACI-SE-E[4], RACI-SE-U[4], RACI-NL[4], RACI-ES[4], RACI-i2b2, ReAct, Dutch (including AMC, BeSt, LUMC and DREAM), anti-TNF response to therapy collection (ACR-REF: BRAGGSS, BRAGGSS2, ERA, KI and TEAR), CORRONA, Vanderbilt, three studies from the GARNET consortium (BioBank Japan Project[3], Kyoto University[3] and IORRA[3]), and Korea. Of these, GWAS data of 4,309 RA cases and 8,700 controls from six studies (RACI-i2b2, ReAct, Dutch, ACR-REF, CORRONA and Vanderbilt) have not been previously published.

The 3,708 RA cases and 5,535 controls enrolled in the stage 2 *in silico* replication study were obtained from two studies of Europeans (2,780 RA cases and 4,700 controls from Genentech and SLEGEN) and Asians (928 RA cases and 835 controls from China) (H.X. *et al.*, manuscript submitted). The 6,938 RA cases and 6,658 controls enrolled in the stage 3 *de novo* replication study were obtained from two studies of Europeans (995 RA cases and 1,101 controls from CANADAII[2]) and Asians (5,943 RA cases and 5,557 controls from BioBank Japan Project, Kyoto University and IORRA[3]).

All subjects in the stage 1, stage 2 and stage 3 studies were confirmed to be independent through analysis of overlapping SNP markers. Any duplicate subjects were removed from the stage 2 and stage 3 replication studies, leading to slightly different sample sizes compared with previous studies that used these same collections[2,3].

All participants provided written informed consent for participation in the study as approved by the ethical committees of each of the institutional review boards. Detailed descriptions of the study design, participating cohorts and the clinical characteristics of the RA cases are provided in detail in Extended Data Fig. 1 and Extended Data Table 1a, as well as in previous reports[2–4].

Genotyping, quality control and genotype imputation of GWAS data. Genotyping platforms and quality control criteria of GWAS, including cut-off values for sample call rate, SNP call rate, minor allele frequency (MAF), and Hardy–Weinberg equilibrium (HWE) *P* value, covariates in the analysis, and imputation reference panel information are provided for each study in Extended Data Table 1b. All studies were analysed based on the same analytical protocol, including exclusion of closely related subjects and outliers in terms of ancestries, as described elsewhere[3]. After applying quality control criteria, whole-genome genotype imputation was performed using 1000 Genomes Project Phase I (α) European ($n = 381$) and Asian ($n = 286$) data as references[11]. We excluded monomorphic or singleton SNPs or SNPs with deviation of HWE ($P < 1.0 \times 10^{-7}$) from each of the reference panels. GWAS data were split into ~300 chunks that evenly covered whole-genome regions and additionally included 300 kb of duplicated regions between neighbouring chunks. Immunochip data were split into ~2,000 chunks that included each of the targeted regions or SNPs on the array. Each chunk was pre-phased and imputed by using minimac (release stamp 2011-10-27). SNPs in the X chromosome were imputed for males and females separately. We excluded imputed SNPs that were duplicated between chunks, SNPs with MAF < 0.005 in RA cases or controls, or with low imputation score (Rsq < 0.5 for genome-wide array and < 0.7 for Immunochip) from each study. We found that imputation of Immunochip effectively increased the number of the available SNPs by 7.0 fold (from ~129,000 SNPs to ~924,000 SNPs) to cover ~12% of common SNPs (MAF > 0.05) included in the 1000 Genomes Project reference panel for European ancestry[11].

Stage 1 trans-ethnic genome-wide meta-analysis. Associations of SNPs with RA were evaluated by logistic regression models assuming additive effects of the allele dosages including top 5 or 10 principal components as covariates (if available) using mach2dat v.1.0.16 (Extended Data Table 1b). Allele dosages of the SNPs in X chromosome were assigned as 0/1/2 for females and 0/2 for males and analysed separately. Meta-analysis was performed for the trans-ethnic study (both Europeans and Asians), European study, and Asian study separately. The SNPs available in ≥3 studies were evaluated in each GWAS meta-analysis, which yielded ~10 million autosomal and X-chromosomal SNPs. Information about the SNPs, including the coded alleles, was oriented to the forward strand of the NCBI build 37 reference sequence. Meta-analysis was conducted by an inverse-variance method assuming a fixed-effects model on the effect estimates (β) and the standard errors of the allele dosages using the Java source code implemented by the authors[25]. Double GC correction was carried out using the inflation factor (λ_{GC}) obtained from the results of each GWAS and the GWAS meta-analysis[25] after removing the SNPs located ± 1 Mb from known RA loci or in the MHC region (chromosome 6, 25–35 Mb). Although there is not yet uniform consensus on the application of double GC correction, we note that potential effects of double GC correction would not be substantial in our study because of the small values of the inflation factors in the GWAS meta-analysis ($\lambda_{GC} < 1.075$ and λ_{GC} adjusted for 1,000 cases and 1,000 controls ($\lambda_{GC\_1,000}$) < 1.005; Extended Data Table 1b).

As for the definition of known RA risk loci in this study, we included the loci that showed significant associations in one of the previous studies ($P < 5.0 \times 10^{-8}$) or that had been replicated in independent cohorts. We consider the locus including multiple independent signals of associations as a single locus, such as the MHC locus[12] and *TNFAIP3* (ref. 4). Although 6 of these 59 loci previously identified as known RA risk loci did not reach a suggestive level of association (defined as $P < 5.0 \times 10^{-6}$) in our stage 1 meta-analysis, previous studies have gone on to replicate most of these associations in additional samples (Supplementary Table 1)[2,3]. Thus, the number of confirmed RA risk loci is 101 (including the MHC region).

Stage 2 and stage 3 replication studies. *In silico* (stage 2) and *de novo* (stage 3) replication studies were conducted using independent European and Asian subjects (Extended Data Table 1). The 146 loci that satisfied $P < 5.0 \times 10^{-6}$ in the stage 1 trans-ethnic, European or Asian GWAS meta-analysis were selected for the stage 2 *in silico* replication study. The SNPs that demonstrated the most significant associations were selected from each of the loci. When the SNP was not available in replication data sets, a proxy SNP with the highest linkage disequilibrium ($r^2 > 0.80$) was alternatively assessed. GWAS quality control, genotype imputation and association analysis were assessed in the same manner as in the stage 1 GWAS. For the 60 loci that demonstrated suggestive associations in the combined results of the stage 1 GWAS meta-analysis and the stage 2 *in silico* replication study but were not included as a known RA risk locus, we calculated statistical power to newly achieve a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ for Europeans and Asians separately, which were estimated based on the allele frequencies, ORs and *de novo* replication sample sizes of the populations. We then selected the top 20 SNPs with the highest statistical power for Europeans and Asians separately (in total 32 SNPs), and conducted the stage 3 *de novo* replication study. Genotyping methods, quality control and confirmation of subject independence in the stage 3 *de novo* replication study were described previously[2,3]. The combined study of the stage 1 GWAS meta-analysis and the stages 2 and 3 replication studies was conducted by an inverse-variance method assuming a fixed-effects model[25].

Trans-ethnic and functional annotations of RA risk SNPs. Trans-ethnic comparisons of RAF (in the reference panels), ORs and explained heritability were conducted using the results of the stage 1 GWAS meta-analysis of Europeans and Asians. Correlations of RAF and OR were evaluated using Spearman's correlation test. ORs were defined based on minor alleles in Europeans. Explained heritability was estimated by applying a liability-threshold model assuming disease prevalence of 0.5% (ref. 10) and using the RAF and OR of the population(s) according to the genetic risk model. For the population-specific genetic risk model, the RAF and OR of the same population was used. For the trans-ethnic genetic risk model, the RAF of the population but the OR of the other population was used.

Details of the overlap enrichment analysis of RA risk SNPs with H3K4me3 peaks have been described elsewhere[13]. Briefly, we evaluated whether the RA risk SNPs (outside of the MHC region) and SNPs in linkage disequilibrium ($r^2 > 0.80$) with them were enriched in overlap with H3K4me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) assay peaks of 34 cell types obtained from the National Institutes of Health Roadmap Epigenomics Mapping Consortium, by a permutation procedure with $\times 10^5$ iterations.

Fine mapping of causal risk alleles. For fine mapping of the causal risk alleles, we selected the 31 RA risk loci where the risk SNPs yielded $P < 1.0 \times 10^{-3}$ in the stage 1 GWAS meta-analysis of both Europeans and Asians with the same directional effects of alleles (outside of the MHC region). For fine mapping using linkage-disequilibrium structure differences between the populations, we calculated average numbers of the SNPs in linkage disequilbrium ($r^2 > 0.80$) in Europeans, Asians, and in both Europeans and Asians, separately.

For fine mapping using H3K4me3 peaks of T_{reg} primary cells, we first evaluated H3K4me3 peak overlap enrichment of the SNPs in linkage disequilbrium (in Europeans and Asians) compared with the neighbouring SNPs (±2 Mb). We fixed the SNP positions but physically slid H3K4me3 peak positions by 1 kb bins within ±2 Mb regions of the risk SNPs, and calculated overlap of the SNPs in linkage disequilibrium with H3K4me3 peaks for each sliding step, and evaluated the significance of overlap in the original peak positions by a one-sided exact test assuming enrichment of overlap. For the 10 loci that demonstrated significant overlap ($P < 0.05$), we calculated the average number of the SNPs that were in linkage disequilibrium in both Europeans and Asians and also included in H3K4me3 peaks.

Pleiotropy analysis. We downloaded phenotype-associated SNPs and phenotype information from the National Human Genome Research Institute (NHGRI) GWAS catalogue database[26] on 31 January, 2013. We selected 4,676 significantly associated SNPs ($P < 5.0 \times 10^{-8}$) corresponding to 311 phenotypes (other than RA). We manually curated the phenotypes by combining the same but differently named phenotypes into a single phenotype (for example, from 'urate levels', 'uric acid levels' and 'renal function-related traits (urea)' to 'urate levels'), or splitting merged phenotypes into sub-categorical phenotypes (for example, from 'white blood cell types' into 'neutrophil counts', 'lymphocyte counts', 'monocyte counts', 'eosinophil counts' or 'basophil counts'). Lists of curated phenotypes and SNPs are available at http://plaza.umin.ac.jp/~yokada/datasource/software.htm.

For each of the selected NHGRI GWAS catalogue SNPs and the RA risk SNPs identified by our study (located outside of the MHC region), we defined the genetic region based on ±25 kb of the SNP or the neighbouring SNP positions in moderate linkage disequilibrium with it in Europeans or Asians ($r^2 > 0.50$). If multiple different SNPs with overlapping regions were registered for the same phenotype, they were merged into a single region. We defined 'region-based pleiotropy' as two phenotype-associated SNPs sharing part of their genetic regions or sharing any UCSC hg19 reference gene(s) that partly overlapped each of the regions (Extended Data Fig. 4a). We defined 'allele-based pleiotropy' as two phenotype-associated SNPs that were in linkage disequilibrium in Europeans or Asians ($r^2 > 0.80$). We defined the direction of an effect as 'concordant' with RA risk if the RA risk allele also leads to increased risk of the disease or increased dosage of the quantitative trait; similarly, we defined relationships as 'discordant' if the RA risk allele is associated with decreased risk of the disease phenotype (or if the RA risk allele leads to decreased dosage of the quantitative trait).

We evaluated statistical significance of region-based pleiotropy of the registered phenotypes with RA by a permutation procedure with $\times 10^7$ iterations. When one phenotype had n loci of which m loci were in region-based pleiotropy with RA, we obtained a null distribution of m by randomly selecting n SNPs from obtained NHGRI GWAS catalogue data and calculating the number of the observed region-based pleiotropy with RA for each of the iteration steps. For estimation of the null distribution, we did not include the SNPs associated with several autoimmune diseases that were previously reported to share pleiotropic associations with RA (Crohn's disease, type 1 diabetes, multiple sclerosis, coeliac disease, systemic lupus erythematosus, ulcerative colitis and psoriasis)[2].

Prioritization of biological candidate genes from RA risk loci. For RA risk SNPs outside of the MHC region, functional annotations were conducted by Annovar (hg19). RA risk SNPs were classified if any of the SNPs in linkage disequilibrium ($r^2 > 0.80$) in Europeans or Asians were annotated in order of priority of missense (or nonsense), synonymous or non-coding (with or without *cis*-eQTL) SNPs. We also applied this SNP annotation scheme to 10,000 randomly selected genome-wide common SNPs (MAF > 0.05 in Europeans or Asians).

We then assessed *cis*-eQTL effects by referring two eQTL data sets: the study for peripheral blood mononuclear cells (PBMCs) obtained from 5,311 European subjects[6] and newly generated cell-specific eQTL analysis for $CD4^+$ T cells and $CD14^+CD16^-$ monocytes from 212 European subjects (ImmVar project; T.R. *et al.*, manuscript submitted). When the RA risk SNP was not available in eQTL data sets, we alternatively used the results of best proxy SNPs in linkage disequilibrium with the highest r^2 value (>0.80). We applied the significance thresholds defined in the original studies (FDR $q < 0.05$ for PBMC eQTL and gene-based permutation $P < 0.05$ for cell-specific eQTL).

We obtained PID genes and their classification categories as defined by the International Union of Immunological Societies Expert Committee[14], downloaded cancer somatic mutation genes from the Catalogue of Somatic Mutations in Cancer (COSMIC) database[15], and downloaded knockout mouse phenotype labels and gene information from the Mouse Genome Informatics (MGI) database[16] on 31 January, 2013 (Supplementary Tables 2–5). We defined 377 RA risk genes included in the 100 RA risk loci (outside of the MHC region) according to the criteria described in the previous section (±25 kb or $r^2 > 0.50$), and evaluated overlap with PID categories, cancer phenotypes with registered somatic mutations, and phenotype labels of knockout mouse genes with human orthologues. Statistical significance of enrichment in gene overlap was assessed by a permutation procedure with $\times 10^6$ iterations. For each iteration step, we randomly selected 100 genetic loci matched for number of nearby genes with those in non-MHC 100 RA risk loci. When one gene category had m genes overlapping with RA risk genes, we obtained a null distribution of m by calculating the number of genes in the selected loci overlapping with RA risk genes for each iteration step.

We conducted molecular pathway enrichment analysis using MAGENTA software[9] and adopting Ingenuity and BIOCARTA databases as pathway information resources. We conducted two patterns of analyses by inputting genome-wide SNP P values of the current trans-ethnic meta-analysis (stage 1) and the previous meta-analysis of RA[2] separately. As the previous meta-analysis was conducted using imputed data based on HapMap Phase II panels, we re-performed the meta-analysis using the same subjects but with newly imputed genotype data based on the 1000 Genomes Project reference panel[11] to make SNP coverage conditions identical between the meta-analyses. Significance of the molecular pathway was evaluated by FDR q values obtained from $\times 10^5$ iterations of permutations.

We scored each of the genes included in the RA risk loci (outside of the MHC region) by adopting the following eight selection criteria and calculating the number of the satisfied criteria: (1) genes for which RA risk SNPs or any of the SNPs in linkage disequilibrium ($r^2 > 0.80$) with them were annotated as missense variants; (2) genes for which significant *cis*-eQTL of any of PBMCs, T cells or monocytes were observed for RA risk SNPs (FDR $q < 0.05$ for PBMCs and permutation $P < 0.05$ for T cells and monocytes); (3) genes prioritized by PubMed text mining using GRAIL[7] with gene-based $P < 0.05$; (4) genes prioritized by PPI network using DAPPLE[8] with gene-based $P < 0.05$; (5) PID genes[14]; (6) haematological cancer somatic mutation genes[15]; (7) genes for which ≥2 of associated phenotype labels ('haematopoietic system phenotype', 'immune system phenotype' and 'cellular phenotype'; $P < 1.0 \times 10^{-4}$) were observed for knockout mouse[16]; and (8) genes prioritized by molecular pathway analysis using MAGENTA[9], which were included in the significantly enriched pathways (FDR $q < 0.05$) with gene-based $P < 0.05$. Because these criteria showed weak correlations with each other ($R^2 < 0.26$; Extended Data Fig. 6c), each gene was given a score based on the number of criteria that were met (scores ranging from 0–8 for each gene). We defined the genes with a score ≥2 as 'biological RA risk genes'.

For each gene in RA risk loci, we evaluated whether the gene was the nearest gene to the RA risk SNP within the risk locus, or whether the RA risk SNP (or SNPs in linkage disequilibrium with it) of the gene overlapped with H3K4me3 histone peaks of cell types. The difference in proportions of genes that were the nearest gene to biological RA risk genes (score ≥2) and non-biological genes (score <2) was evaluated by using Fisher's exact test implemented in R statistical software (v.2.15.2). The difference in the proportions of genes overlapping with T_{reg} primary cell H3K4me3 peaks between biological and non-biological genes was assessed by a permutation procedure by shuffling the overlapping status of RA risk SNPs/loci with $\times 10^5$ iterations.

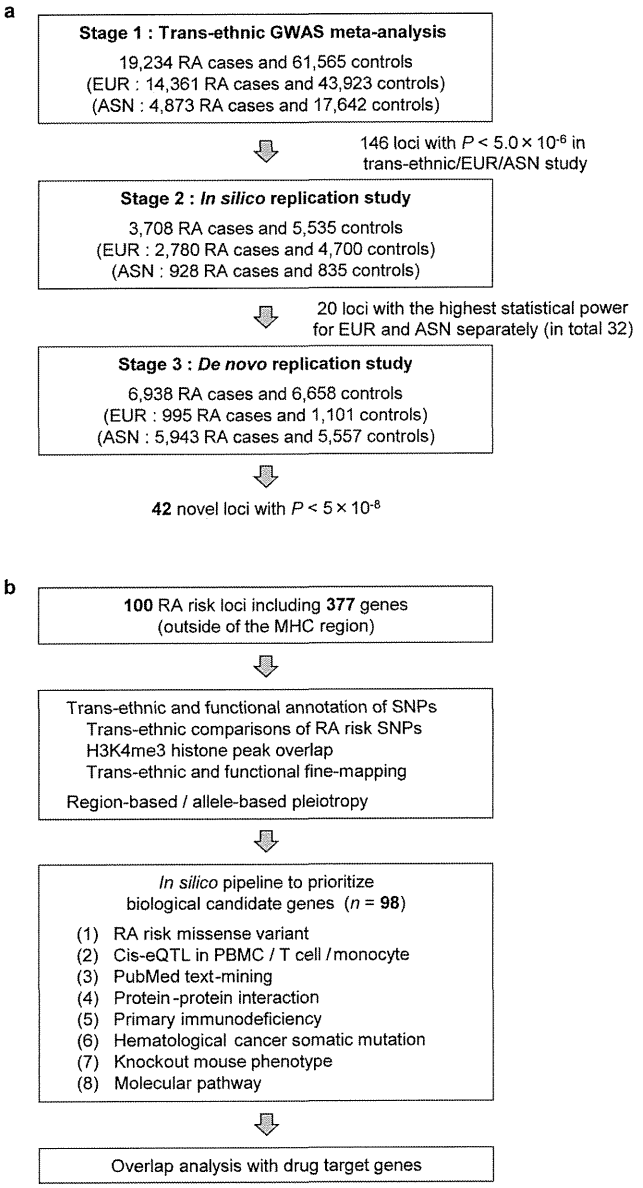
Drug target gene enrichment analysis. We obtained drug target genes and corresponding drug information from DrugBank[17] and the Therapeutic Targets Database (TTD)[18] on 31 January, 2013, as well as additional literature searches. We selected drug target genes that had pharmacological activities (for the genes from DrugBank) and human orthologues, and that were annotated to any of the approved, clinical trial or experimental drugs (Supplementary Table 6). We manually extracted drug target genes annotated to approved RA drugs on the basis of discussions with professional rheumatologists (Extended Data Fig. 7a). We extracted genes in direct PPI with biological RA risk genes by using the InWeb database[27]. To take account of potential dependence between PPI genes and drug target genes, overlap of biological RA risk genes and genes in direct PPI with them with drug target genes was assessed by a permutation procedure with $\times 10^5$ iterations.

Let x be the set of the biological RA risk genes and genes in direct PPI with them (n_x genes), y be the set of genes with protein products that are the direct target of approved RA drugs (n_y genes), and z be the set of genes with protein products that are the direct target of all approved drugs (n_z genes). We defined $n_{x\cap y}$ and $n_{x\cap z}$ as the numbers of genes overlapping between x and y and between x and z, respectively. For each of 10,000 iteration steps, we randomly selected a gene set of x' including n_x genes from the entire PPI network (12,735 genes). We defined $n_{x\cap y}'$ and $n_{x\cap z}'$ as the numbers of genes overlapping between x' and y, and between x' and z, respectively. The distributions of $n_{x\cap y}'$, $n_{x\cap z}'$ and $n_{x\cap y}'/n_{x\cap z}'$ obtained from the total iterations were defined as the null distributions of $n_{x\cap y}$, $n_{x\cap z}$, and $n_{x\cap y}/n_{x\cap z}$, respectively. Fold enrichment of overlap with approved RA drug target genes was defined as $n_{x\cap y}/m(n_{x\cap y}')$, where $m(t)$ represents the mean value of the distribution of t. Fold enrichment of overlap with approved all drug target genes was defined as $n_{x\cap z}/m(n_{x\cap z}')$. Relative fold enrichment of overlap with RA drug target genes and with all drug target genes was defined as $(n_{x\cap y}/n_{x\cap z})/m(n_{x\cap y}'/n_{x\cap z}')$. Significance of the enrichment was evaluated by one-sided permutation tests examining $n_{x\cap y}$, $n_{x\cap z}$, and $n_{x\cap y}/n_{x\cap z}$ in their null distributions.

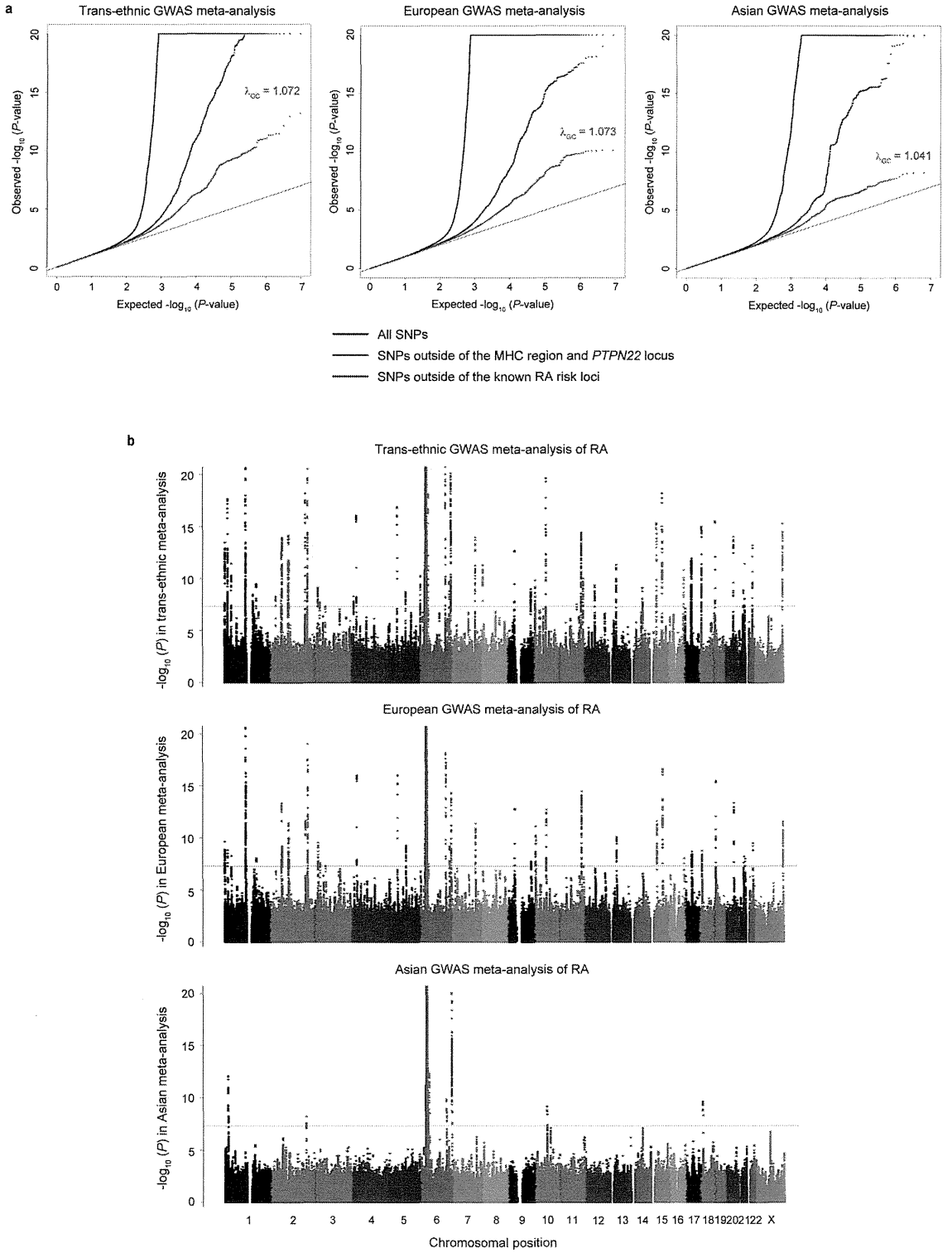
Web resources. The following websites provide valuable additional resources. Summary statistics from the GWAS meta-analysis, source codes, and data sources have been deposited at http://plaza.umin.ac.jp/~yokada/datasource/software.htm; GARNET consortium, http://www.twmu.ac.jp/IOR/garnet/home.html; i2b2, https://www.i2b2.org/index.html; SLEGEN, http://www.lupusresearch.org/lupus-research/slegen.html; 1000 Genomes Project, http://www.1000genomes.org/; minimac, http://genome.sph.umich.edu/wiki/Minimac; mach2dat, http://www.sph.umich.edu/csg/abecasis/MACH/index.html; Annovar, http://www.openbioinformatics.org/annovar/; ImmVar, http://www.immvar.org/; NIH Roadmap Epigenomics Mapping Consortium, http://www.roadmapepigenomics.org/; NHGRI GWAS catalogue, http://www.genome.gov/GWAStudies/; COSMIC, http://cancer.sanger.ac.uk/cancergenome/projects/

cosmic/; MGI, http://www.informatics.jax.org/; MAGENTA, http://www.broadinstitute.org/mpg/magenta/; Ingenuity, http://www.ingenuity.com/; BIOCARTA, http://www.biocarta.com/; GRAIL, http://www.broadinstitute.org/mpg/grail/; DAPPLE, http://www.broadinstitute.org/mpg/dapple/dapple.php; R statistical software, http://www.r-project.org/; DrugBank, http://www.drugbank.ca/; TTD, http://bidd.nus.edu.sg/group/ttd/ttd.asp.

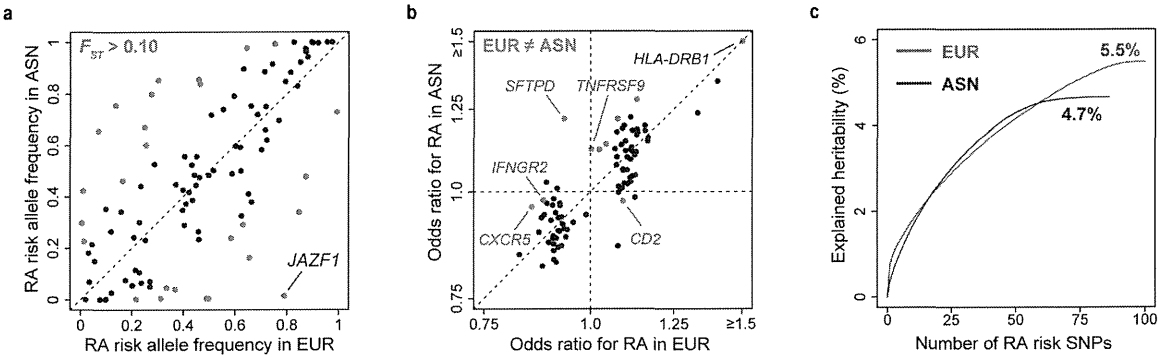
24. Arnett, F. C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31,** 315–324 (1988).
25. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nature Genet.* **44,** 904–909 (2012).
26. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
27. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnol.* **25,** 309–316 (2007).
28. Ueda, H. *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423,** 506–511 (2003).
29. Elliott, P. *et al.* Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *J. Am. Med. Assoc.* **302,** 37–48 (2009).
30. Cortes, A. *et al.* Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nature Genet.* **45,** 730–738 (2013).



Extended Data Figure 1 | An overview of the study design. **a**, We conducted a three-stage trans-ethnic meta-analysis in total of 29,880 RA cases and 73,758 controls of European (EUR) and Asian (ASN) ancestry. The stage 1 GWAS meta-analysis included 19,234 RA cases and 61,565 controls from 22 studies, which was followed by the stage 2 *in silico* replication study (3,708 RA cases and 5,535 controls) and stage 3 *de novo* replication study (6,938 RA cases and 6,658 controls). In the combined study of stages 1–3, we identified 42 novel RA risk loci, which increased the total number of RA risk loci to 101. **b**, Using the 100 RA risk loci (outside of the MHC region), we conducted trans-ethnic and functional annotation of the RA risk SNPs. We constructed an *in silico* bioinformatics pipeline to prioritize biological candidate genes. We adopted eight criteria to score each of 377 genes in the RA risk loci: (1) RA risk missense variant; (2) *cis*-eQTL; (3) PubMed text mining; (4) PPI; (5) PID; (6) haematological cancer somatic mutation; (7) knockout mouse phenotype; and (8) molecular pathway. Our study also demonstrated that these biological candidate genes in RA risk loci are significantly enriched in overlap with target genes for approved RA drugs.



Extended Data Figure 2 | Quantile–quantile plots and Manhattan plots of *P* values in the GWAS meta-analysis. **a**, Quantile–quantile plots of *P* values in the stage 1 GWAS meta-analysis for trans-ethnic, European and Asian ancestries. The *x*-axis indicates the expected $-\log_{10}$ (*P* values). The *y*-axis indicates the observed $-\log_{10}$ (*P* values) after the application of double GC correction. The SNPs for which observed *P* values were less than 1.0×10^{-20} are indicated at the upper limit of each plot. Black, blue and red dots represent the association results of all SNPs, SNPs outside of the MHC region and *PTPN22* locus, and SNPs outside of the known RA risk loci, respectively. Double GC correction was applied based on the inflation factor, λ_{GC}, which was estimated from the SNPs outside of the known RA loci and indicated in each plot. **b**, Manhattan plots of *P* values in the stage 1 GWAS meta-analysis for trans-ethnic, European and Asian ancestries. The *y*-axis indicates the $-\log_{10}$ (*P* values) of genome-wide SNPs in each GWAS meta-analysis. The horizontal grey line represents the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$. The SNPs for which *P* values were less than 1.0×10^{-20} are indicated at the upper limit of each plot.



d

Cell types	*P* for H3K4me3 enrichment
T_{reg} primary cells	$\leq 1.0\times10^{-5}$
$CD4^+$ memory primary cells	3.0×10^{-5}
$CD4^+$ naive primary cells	0.0041
$CD8^+$ memory primary cells	0.0065
Smooth muscle, rectal	0.034
Mucosa, colon	0.038
$CD8^+$ naive primary cells	0.12
Mucosa, stomach	0.13
$CD34^+$ primary cells	0.18
$CD34^+$ cultured cells	0.19
Mobilized $CD34^+$ primary cells	0.19
$CD19^+$ primary cells	0.24
$CD3^+$ primary cells	0.30
Mucosa, duodenum	0.40
Muscle satellite cultured cells	0.46
Cingulate gyrus (brain)	0.53
Skeletal muscle	0.77
Mucosa, rectal	0.77
Smooth muscle, colon	0.79
Mesenchymal stem cells (adipose)	0.81
Adipose nuclei	0.84
Smooth muscle, duodenum	0.85
Mid frontal lobe (brain)	0.86
Hippocampus middle (brain)	0.91
Mesenchymal stem cells (bone marrow)	0.91
Pancreatic islets	0.93
Inferior temporal lobe (brain)	0.93
Substantia nigra (brain)	0.93
Adult kidney	0.94
Adult liver	0.95
Mesenchymal stem cells (adipocyte)	0.98
Mesenchymal stem cells (chondrocytes)	0.99
Anterior caudate (brain)	0.99
Smooth muscle, stomach	0.99

e

31 RA *risk loci with* $P < 10^{-3}$
in EUR and ASN meta-analysis

21.9 SNPs/locus in LD (r^2>0.8) in EUR

37.3 SNPs/locus in LD (r^2>0.8) in ASN

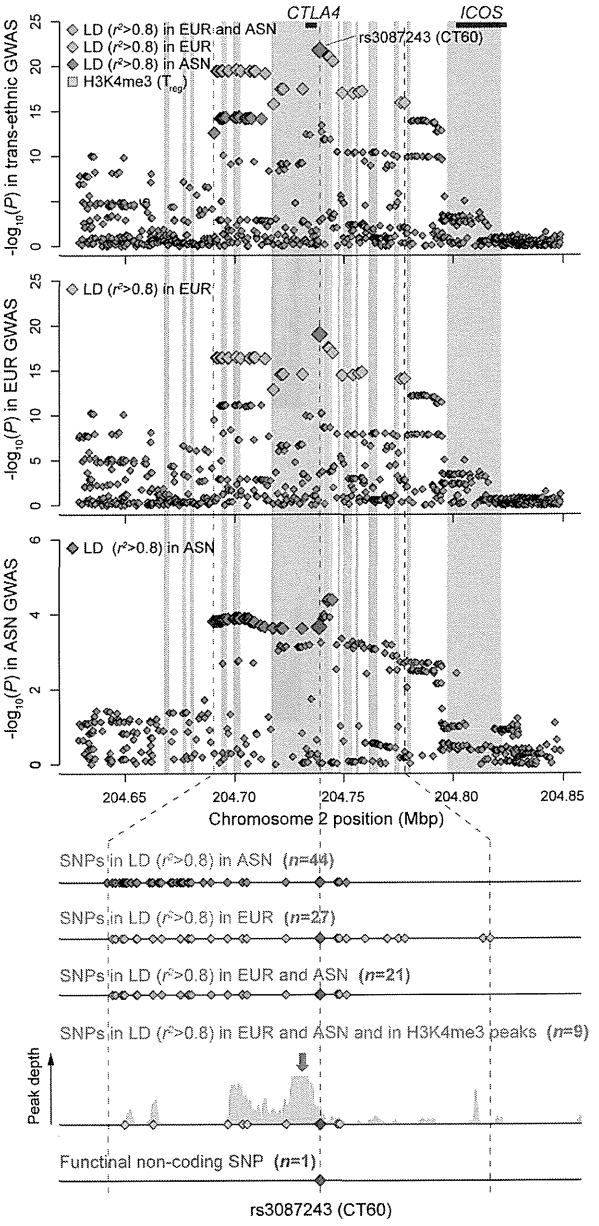
15.0 SNPs/locus in LD (r^2>0.8) in EUR and ASN

10 RA risk loci significantly enriched with
T_{reg} primary cell H3K4me3 peaks (P < 0.05)

10.4 SNPs/locus in LD (r^2>0.8) in EUR and ASN

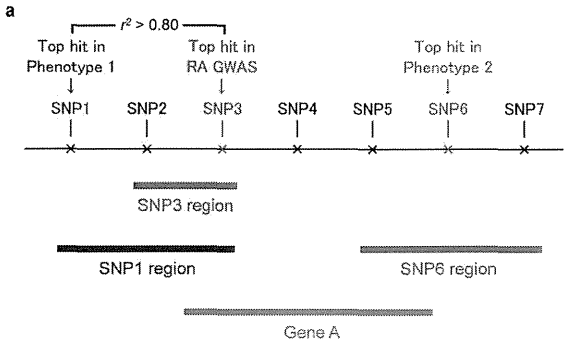
5.9 SNPs/locus in LD (r^2>0.8) in EUR and ASN
and overlapping with H3K4me3 peaks

f



Extended Data Figure 3 | Trans-ethnic and functional annotation of RA risk SNPs. a, b, Comparisons of RAF and OR values between individuals of European (EUR) and Asian (ASN) ancestry from the stage 1 GWAS meta-analysis. ORs were defined based on minor alleles in Europeans. SNPs with $F_{ST} > 0.10$ or SNPs in which the 95% CI of the OR did not overlap between Europeans and Asians are coloured. OR of the SNP in the *HLA-DRB1* locus (≥ 1.5) is plotted at the upper limits of the *x*- and *y*-axes. Five loci demonstrated population-specific associations ($P < 5.0 \times 10^{-8}$ in one population but $P > 0.05$ in the other population without overlap of the 95% CI of the OR) are highlighted by red labels (rs227163 at *TNFRSF9*, rs624988 at *CD2*, rs726288 at *SFTPD*, rs10790268 at *CXCR5* and rs73194058 at *IFNGR2*). **c**, Cumulative curve of explained heritability in each population. **d**, Enrichment analysis for overlap of RA risk SNPs with H3K4me3 peaks in cell types. The most significant cell type is T_{reg} primary cells. **e**, Number of SNPs in the process of trans-ethnic and functional fine mapping. For 31 loci in which the risk SNPs yielded $P < 1.0 \times 10^{-3}$ in both populations (stage 1 GWAS), the number of candidate causal variants was reduced by 40–70% when confined by SNPs in linkage disequilibrium with the RA risk SNPs ($r^2 > 0.80$) in both populations (on average, from 21.9 or 37.3 SNPs in linkage disequiliberium in Europeans or Asians, to 15.0 SNPs in linkage disequilibrium in both populations). Further, for 10 loci in which candidate causal variants significantly overlapped with H3K4me3 peaks in T_{reg} cells ($P < 0.05$), the average number of SNPs was further reduced by half again, from 10.4 to 5.9. **f**, Fine mapping in the *CTLA4* locus, where the functional non-coding variant of CT60 (rs3087243)[28] showed the most significant association with RA. The top three panels indicate regional SNP associations of the locus in the stage 1 GWAS meta-analysis for trans-ethnic, European and Asian ancestries, respectively. The bottom panel indicates the change in the number of the candidate causal variants in each process of fine mapping. Trans-ethnic fine mapping of candidate causal variants decreased the number of candidate variants from 44 (linkage disequilibrium in Asians) and 27 (linkage disequilibrium in Europeans) to 21 (linkage disequilibrium in both populations). As these SNPs were significantly enriched in overlap with H3K4me3 peaks in T_{reg} cells compared with the surrounding SNPs ($P = 0.037$), we confined the candidate variants into nine by additionally selecting the SNPs included in H3K4me3 peaks. CT60 was included in these finally selected nine SNPs, and also located at the vicinity of a H3K4me3 peak summit (indicated by a red arrow).

a



RA and Phenotype 1 : Both region-based and allele-based pleiotropy.

RA and Phenotype 2 : Region-based pleiotropy only.

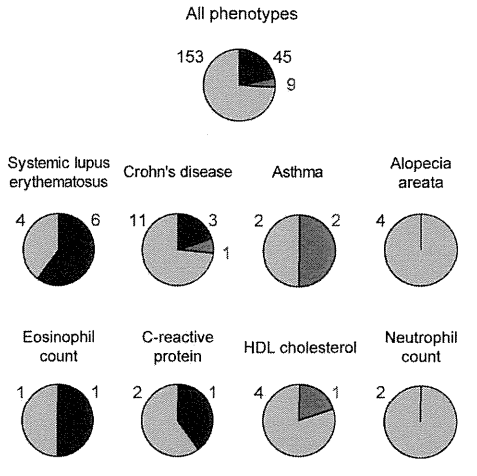
b

Phenotype in GWAS catalogue	No. loci	Region-based pleiotropy		Allele-based pleiotropy
		No. overlap	P-value	
Type 1 diabetes	42	15	$<1.0\times10^{-7}$	7
Crohn's disease	79	15	$<1.0\times10^{-7}$	4
Systemic lupus erythematosus	22	10	$<1.0\times10^{-7}$	6
Celiac disease	26	10	$<1.0\times10^{-7}$	3
Vitiligo	23	9	$<1.0\times10^{-7}$	3
Primary biliary cirrhosis	22	7	2.4×10^{-6}	3
Alopecia areata	5	4	4.5×10^{-6}	0
Ulcerative colitis	52	9	2.5×10^{-5}	3
Multiple sclerosis	52	9	2.5×10^{-5}	2
Chronic lymphocytic leukemia	9	4	9.1×10^{-5}	0
Kawasaki disease	5	3	2.4×10^{-4}	2
Graves' disease	5	3	2.4×10^{-4}	1
Systemic sclerosis	5	3	2.4×10^{-4}	1
Fibrinogen	8	3	0.0012	1
Asthma	17	4	0.0015	2
Psoriasis	18	4	0.0019	1
Hypothyroidism	4	2	0.0041	2
Basal cell carcinoma	5	2	0.0069	0
Neutrophil count	5	2	0.0069	0
HDL cholesterol	46	5	0.014	1
Eosinophil counts	8	2	0.018	1
C-reactive protein	20	3	0.020	1
Melanoma	11	2	0.034	0
Myasthenia gravis	2	1	0.039	1
Primary sclerosing cholangitis	2	1	0.039	0
Soluble ICAM-1	2	1	0.039	0

c

SNP	Chr.	Position (bp)	A1/A2	Gene	Phenotype	Direction
chr1:2523811	1	2,523,811	G/A	*TNFRSF14-MMEL1*	Multiple sclerosis	Concordant
					Hypothyroidism	Concordant
rs2476601	1	114,377,568	A/G	*PTPN22*	Myasthenia gravis	Concordant
					Crohn's disease	Discordant
					Type 1 diabetes	Concordant
					C-reactive protein	Concordant
rs2228145	1	154,426,970	A/C	*IL6R*	Asthma	Discordant
					sIL-6R	Discordant
					Fibrinogen	Concordant
rs2317230	1	157,674,997	T/G	*FCRL3*	Graves' disease	Concordant
rs34695944	2	61,124,850	C/T	*REL*	Hodgkin lymphoma	Concordant
					Psoriasis	Discordant
rs11889341	2	191,943,742	T/C	*STAT4*	Systemic sclerosis	Concordant
					Systemic lupus erythematosus	Concordant
rs3087243	2	204,738,919	G/A	*CTLA4*	Type 1 diabetes	Concordant
rs11933540	4	26,120,001	C/T	*C4orf52*	Type 1 diabetes	Concordant
rs17264332	6	138,005,515	G/A	*TNFAIP3*	Celiac disease	Concordant
					Ulcerative colitis	Concordant
rs7752903	6	138,227,364	G/T	*TNFAIP3*	Systemic lupus erythematosus	Concordant
chr7:128580042	7	128,580,042	G/A	*IRF5*	Ulcerative colitis	Concordant
					Systemic lupus erythematosus	Concordant
rs2736337	8	11,341,880	C/T	*BLK*	Kawasaki disease	Concordant
					Systemic lupus erythematosus	Concordant
rs1516971	8	129,542,100	T/C	*PVT1*	Ovarian cancer	Concordant
					Crohn's disease	Concordant
rs947474	10	6,390,450	A/G	*PRKCQ*	Type 1 diabetes	Concordant
rs2671692	10	50,097,819	A/G	*WDFY4*	Systemic lupus erythematosus	Concordant
rs726288	10	81,706,973	T/C	*SFTPD*	Serum SP-D levels	Concordant
rs4409785	11	95,311,422	C/T	*CEP57*	Vitiligo	Concordant
rs10790268	11	118,729,391	G/A	*CXCR5*	Primary biliary cirrhosis	Concordant
rs61432431	11	128,322,622	C/T	*ETS1*	Systemic lupus erythematosus	Concordant
					Polycystic ovary syndrome	Discordant
rs773125	12	56,394,954	A/G	*CDK2*	Vitiligo	Discordant
					Type 1 diabetes	Discordant
					Eosinophil counts	Concordant
					Hypothyroidism	Concordant
					Platelet-related traits	Concordant
					Type 1 diabetes	Concordant
rs10774624	12	111,833,788	G/A	*SH2B3-PTPN11*	Blood pressure and hypertension	Concordant
					Vitiligo	Concordant
					Retinal vascular caliber	Concordant
					CKD	Concordant
					Celiac disease	Concordant
rs1950897	14	68,760,141	T/C	*RAD51B*	Primary biliary cirrhosis	Concordant
rs13330176	16	86,019,087	A/T	*IRF8*	Multiple sclerosis	Concordant
					Primary biliary cirrhosis	Concordant
					Ulcerative colitis	Concordant
chr17:38031857	17	38,031,857	G/T	*IKZF3-CSF3*	Crohn's disease	Concordant
					Asthma	Discordant
					Type 1 diabetes	Concordant
rs4239702	20	44,749,251	C/T	*CD40*	Kawasaki disease	Concordant
rs2236668	21	45,650,009	C/T	*ICOSLG-AIRE*	Celiac disease	Concordant
rs11089637	22	21,979,096	C/T	*UBE2L3-YDJC*	Crohn's disease	Concordant
					HDL	Discordant

d

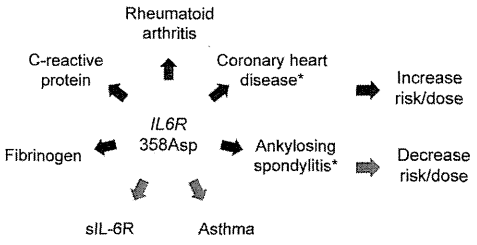


Region- and Allele-based pleiotropy (concordant direction)

Region- and Allele-based pleiotropy (discordant direction)

Region-based pleiotropy only

e

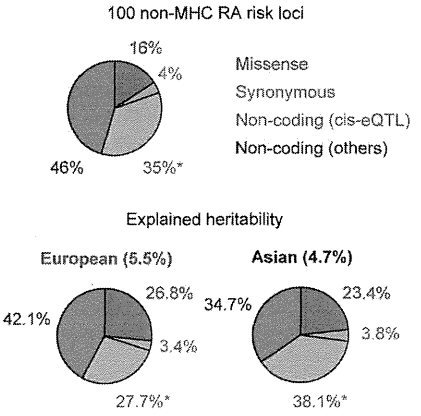


Extended Data Figure 4 | Pleiotropy of RA risk SNPs. **a**, Definition of region-based and allele-based pleiotropy. For each of the RA risk SNPs and SNPs registered in the NHGRI GWAS catalogue (outside of the MHC region), we defined the region on the basis of ±25 kb of the SNP or the neighbouring SNP positions in moderate linkage disequilibrium with it in Europeans or Asians ($r^2 > 0.50$). We defined 'region-based pleiotropy' as two phenotype-associated SNPs sharing part of their genetic regions or any UCSC hg19 reference gene(s) partly overlapping with each of the regions. We defined 'allele-based pleiotropy' as two phenotype-associated SNPs in linkage disequilibrium in Europeans or Asians ($r^2 > 0.80$). **b**, Region-based pleiotropy of the RA risk loci. We found two-thirds of RA risk loci ($n = 66$) demonstrated region-based pleiotropy with other human phenotypes. Phenotypes which showed region-based pleiotropy with RA risk loci are indicated ($P < 0.05$). **c**, Allele-based pleiotropy of the RA risk loci. Allele-based pleiotropy with discordant directional effects to RA risk SNPs are indicated in grey. **d**, Relative proportions of pleiotropic effects (that is, regions and alleles that influence multiple phenotypes) between RA risk loci and 311 phenotypes from the NHGRI GWAS catalogue. Representative examples of disease and biomarker phenotypes are shown. One-quarter of the observed region-based pleiotropic associations (26% = 54/207) were also annotated as having allele-based pleiotropy, although their proportions and directional effects varied among phenotypes. **e**, Allele-based pleiotropy of *IL6R* 358Asp (rs2228145 (A))[5] on multiple disease phenotypes, including increased risk of RA, ankylosing spondylitis and coronary heart disease (asterisks indicate associations obtained from the literature[29,30]) and protection from asthma, as well as levels of biomarkers (increased C-reactive protein (CRP) and fibrinogen but decreased soluble interleukin-6 receptor (sIL6R)).

a

RA risk SNP	r^2	Gene	Missense variants
rs2301888	0.95	*PADI4*	Gly55Ser, Val82Ala, Gly112Ala
rs2476601	1.00	*PTPN22*	Arg620Trp
rs2228145	1.00	*IL6R*	Asp358Ala
rs9826828	0.92	*NCK1*	Ala116Val
	1.00	*NFKBIE*	Val194Ala, Pro175Leu
rs2233424	0.94	*TCTE1*	Arg59His
	0.88	*AARS2*	Val730Met
rs7752903	1.00	*TNFAIP3*	Phe127Cys
rs2671692	0.84	*WDFY4*	Arg1816Gln
rs6479800	0.88	*RTKN2*	Ala288Thr
rs508970	0.90	*CD5*	Ala471Val
rs10774624	0.86	*SH2B3*	Trp262Arg
rs3783782	1.00	*PRKCH*	Val374Ile
rs2582532	1.00	*AHNAK2*	Gly1901Ser
chr17:38031857	0.99	*ZPBP2*	Ser151Ile
	0.99	*GSDMB*	Pro298Ser, Gly291Arg
rs34536443	0.87	*TYK2*	Pro1104Ala
rs2236668	0.94	*ICOSLG*	Trp353Arg
rs5987194	0.96	*IRAK1*	Phe196Ser, Ser453Leu

b



c

	PID classification	No. PID genes	No. overlap with RA genes	Overlap genes	P-value
	All PID genes	194	14	-	1.2×10^{-4}
I	Combined immunodeficiencies	43	3	*PTPRC, RAG1/2, CD40*	0.046
II	Well-defined syndromes	25	2	*ATM, TYK2*	0.12
III	Primary antibody deficiencies	21	2	*CD40, UNG*	0.030
IV	Immune dysregulation	21	4	*CASP8, CASP10, AIRE, IL2RA*	0.0033
V	Phagocyte defects	33	2	*IFNGR2, IRF8*	0.16
VI	Innate immunity	19	0	-	1.0
VII	Autoinflammatory	13	1	*MVK*	0.16
VIII	Complement deficiencies	27	1	*C5*	0.33

d

Cancer type	No. cancer somatic mutation genes	No. overlap with RA genes	Overlap genes	P-value
All cancers	444	23	-	4.7×10^{-5}
Hematological cancers	251	17	-	1.2×10^{-4}
Non-hematological cancers	221	6	-	0.56
Hodgkin lymphoma	10	2	*REL, TNFAIP3*	0.010
B cell non-Hodgkin lymphoma	8	2	*DDX6, FCRL4*	0.015
Non-Hodgkin lymphoma	21	2	*FGFR1OP, HSP90AB1*	0.067
Acute lymphocytic leukemia	29	3	*FCGR2B, AFF3, CDK6*	0.079
Acute myelogenous leukemia	68	2	*ACSL6, PTPN11*	0.47

e

Konckout mouse phenotype category	No. kockout mouse genes with human ortholog	No. overlap with RA genes	P-value
Hematopoietic system phenotype	**2,159**	**86**	$\mathbf{7.0\times10^{-6}}$
Immune system phenotype	**2,622**	**94**	$\mathbf{1.2\times10^{-5}}$
Cellular phenotype	**2,961**	**97**	**0.0015**
Liver/biliary system phenotype	982	35	0.0091
Renal/urinary system phenotype	1,028	35	0.011
Endocrine/exocrine gland phenotype	1,453	45	0.020
Respiratory system phenotype	1,097	31	0.028
Tumorigenesis	807	30	0.049
Normal phenotype	1,599	42	0.18
Homeostasis/metabolism phenotype	3,356	88	0.20
Integument phenotype	1,455	35	0.27
Pigmentation phenotype	355	9	0.31
Cardiovascular system phenotype	1,987	42	0.51
Skeleton phenotype	1,435	34	0.57
Other phenotype	258	6	0.57
No phenotypic analysis	1,053	21	0.59
Mortality/aging	3,952	93	0.75
Adipose tissue phenotype	617	12	0.78
Growth/size phenotype	3,061	67	0.79
Digestive/alimentary phenotype	1,128	22	0.80
Reproductive system phenotype	1,730	37	0.81
Limbs/digits/tail phenotype	748	13	0.82
Taste/olfaction phenotype	123	1	0.85
Hearing/vestibular/ear phenotype	557	8	0.88
Embryogenesis phenotype	1,535	30	0.92
Behavior/neurological phenotype	2,465	46	0.94
Nervous system phenotype	2,805	53	0.95
Craniofacial phenotype	951	15	0.96
Muscle phenotype	1,198	21	0.96
Vision/eye phenotype	1,214	21	0.99

f

Database	Molecular pathway	Pathway enrichment (FDR q) Current study	Previous study
BIOCARTA	B Lymphocyte Cell Surface Molecules	$\mathbf{2.0\times10^{-4}}$	0.26
BIOCARTA	T Cytotoxic Cell Surface Molecules	$\mathbf{3.3\times10^{-4}}$	**0.032**
BIOCARTA	T Helper Cell Surface Molecules	$\mathbf{4.0\times10^{-4}}$	**0.030**
BIOCARTA	Th1/Th2 Differentiation	**0.0025**	**0.0063**
Ingenuity	*IL-10.Signaling*	**0.0026**	0.46
Ingenuity	Interferon.Signaling	**0.0028**	0.13
Ingenuity	GM-CSF.Signaling	**0.0031**	0.43
Ingenuity	T.Cell.Receptor.Signaling	**0.0034**	**0.029**
BIOCARTA	NO2-dependent IL 12 Pathway in NK cells	**0.0044**	0.06
BIOCARTA	IL-22 Soluble Receptor Signaling	**0.0046**	0.39
BIOCARTA	The Co-Stimulatory Signal During T-cell Activation	**0.0046**	0.06
BIOCARTA	Selective expression of chemokine receptors during T-cell polarization	**0.0048**	0.21
Ingenuity	Hepatic.Fibrosis.Hepatic.Stellate.Cell.Activation	**0.0073**	**0.0060**
Ingenuity	p38.MAPK.Signaling	**0.0076**	0.19
Ingenuity	Neuregulin.Signaling	**0.0079**	0.51
Ingenuity	IL-6.Signaling	**0.0082**	0.11
Ingenuity	Glucocorticoid.Receptor.Signaling	**0.0090**	0.18
BIOCARTA	IL-6 signaling	**0.0091**	0.50
BIOCARTA	Influence of Ras and Rho proteins on G1 to S Transition	**0.016**	0.38
BIOCARTA	IL-3 signaling	**0.018**	0.64
BIOCARTA	Adhesion and Diapedesis of Granulocytes	**0.018**	0.15
BIOCARTA	RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage	**0.018**	0.15
Ingenuity	Fc.Epsilon.RI.Signaling	**0.022**	0.19
Ingenuity	JAK.Stat.Signaling	**0.023**	0.48
Ingenuity	IL-2.Signaling	**0.026**	0.17
Ingenuity	PPAR.Signaling	**0.026**	0.24
BIOCARTA	IL-2 Receptor Beta Chain in T cell Activation	**0.027**	0.39
BIOCARTA	Cyclins and Cell Cycle Regulation	**0.028**	0.16
Ingenuity	Leukocyte.Extravasation.Signaling	**0.028**	0.45
BIOCARTA	p53 Signaling Pathway	**0.028**	0.40
BIOCARTA	Role of ERBB2 in Signal Transduction and Oncology	**0.028**	0.51
Ingenuity	B.Cell.Receptor.Signaling	**0.028**	0.45
BIOCARTA	CD40L Signaling	**0.029**	0.16
BIOCARTA	*Cells and Molecules involved in local acute inflammatory response*	**0.034**	0.40
BIOCARTA	Antigen Dependent B Cell Activation	**0.036**	0.06
BIOCARTA	Adhesion and Diapedesis of Lymphocytes	**0.043**	0.60
BIOCARTA	MAPKinase Signaling	**0.044**	0.76
BIOCARTA	Phosphorylation of MEK1 by cdk5/p35 down regulates the MAP kinase	**0.044**	0.59
Ingenuity	NFKB.Signaling	**0.045**	0.05
Ingenuity	*Aryl.Hydrocarbon.Receptor.Signaling*	**0.048**	0.33
Ingenuity	PDGF.Signaling	**0.049**	0.30

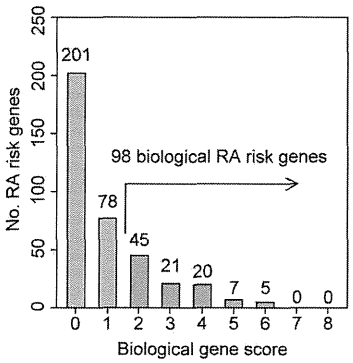
Extended Data Figure 5 | Overlap of RA risk SNPs with biological resources. **a**, Missense variants in linkage disequilibrium ($r^2 > 0.80$ in Europeans or Asians) with RA risk SNPs. When multiple missense variants are in linkage disequilibrium with the RA risk SNP, the highest r^2 value is indicated. **b**, Functional annotation of the SNPs in 100 non-MHC RA risk loci, including the relative proportion of heritability explained by SNP annotations. Although 44% of all RA risk SNPs had *cis*-eQTL, 9 of them overlapped with missense or synonymous variants but 35 of them did not overlap as indicated by asterisks. A list of *cis*-eQTL SNPs and genes can be found in Extended Data Table 2. **c**, Overlap of RA risk genes with human PID and defined categories. **d**, Overlap of RA risk genes with cancer somatic mutation genes. In addition to the categories of all cancers, haematological cancers and non-haematological cancers, cancer types that showed overlap with ≥2 of RA risk genes are indicated. **e**, Overlap of RA risk genes with knockout mouse phenotypes. Knockout mouse phenotypes that satisfied significant enrichment with RA risk genes are indicated in bold ($P < 0.05/30 = 0.0017$). **f**, Molecular pathway analysis of RA GWAS results. Molecular pathways that showed significant enrichment in either the current stage 1 trans-ethnic GWAS meta-analysis or the previous GWAS meta-analysis of RA[2] are indicated in bold (FDR $q < 0.05$).

a

Biological RA risk gene prioritization criteria

(1) RA risk missense variant (n = 19)
(2) Cis-eQTL (n = 51)
(3) PubMed text-mining (n = 90)
(4) Protein-protein interaction (n = 63)
(5) Primary immunodeficiency (n = 15)
(6) Hematological cancer (n = 17)
(7) Knockout mouse phenotype (n = 86)
(8) Molecular pathway (n = 35)

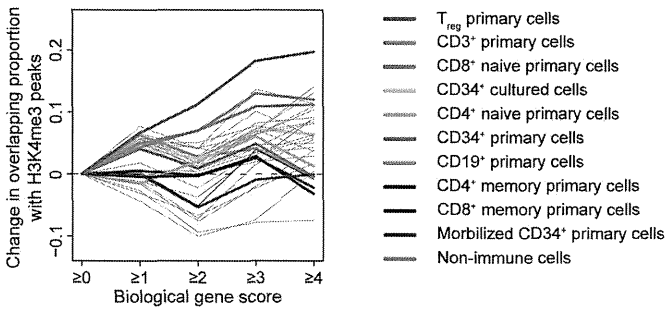
b



c

Correlation of prioritization criteria of biological genes from RA risk loci (R^2)	RA risk missense variant	Cis-eQTL	PubMed text-mining	Protein-protein interaction	Primary immunodeficiency	Hematological cancer	Knockout mouse phenotype	Molecuular pathway
RA risk missense variant	-	0.01	0.03	0.02	0.00	0.00	0.02	0.01
Cis-eQTL	0.01	-	0.05	0.01	0.01	0.00	0.02	0.01
PubMed text-mining	0.03	0.05	-	0.10	0.06	0.03	0.26	0.14
Protein-protein interaction	0.02	0.01	0.10	-	0.04	0.01	0.07	0.06
Primary immunodeficiency	0.00	0.01	0.06	0.04	-	0.00	0.08	0.07
Hematological cancer	0.00	0.00	0.03	0.01	0.00	-	0.03	0.04
Knockout mouse phenotype	0.02	0.02	0.26	0.07	0.08	0.03	-	0.21
Molecular pathway	0.01	0.01	0.14	0.06	0.07	0.04	0.21	-

d

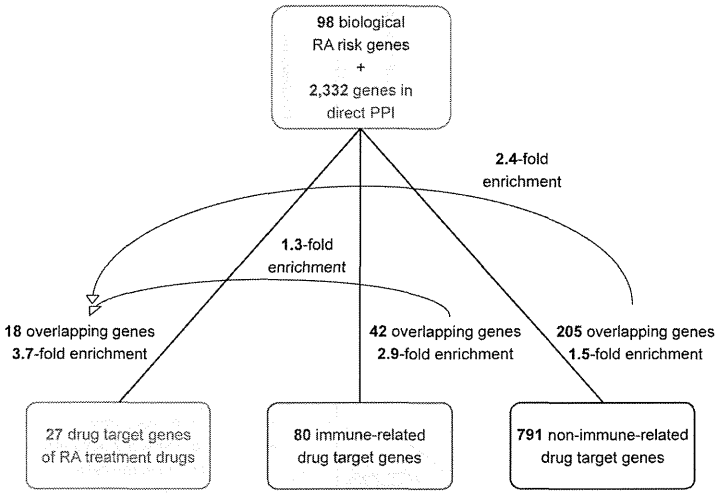


Extended Data Figure 6 | Prioritization of biological candidate genes from RA risk loci. **a**, Prioritization criteria of biological candidate genes from RA risk loci. **b**, Histogram distribution of gene scores. The 98 genes with score ≥2 (orange) were defined as 'biological RA risk genes'. **c**, Correlations of biological candidate gene prioritization criteria. **d**, Change in the overlapping proportions of genes with H3K4me3 peaks by cell type according to score increases. When RA risk SNP of the locus (or SNP in linkage disequilibrium) overlapped with H3K4me3 peaks, genes in the locus were defined as overlapping.

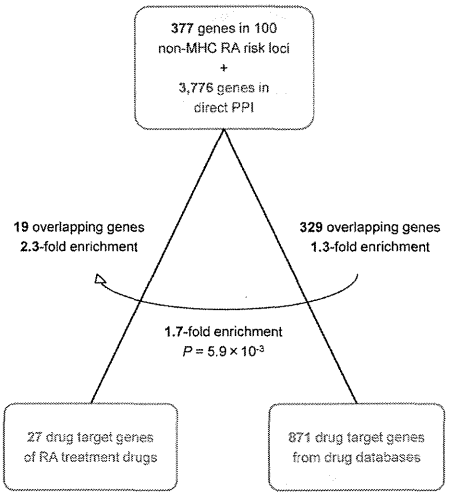
a

RA drug category	Generic name	Target gene
Biologics	Etanercept	*TNF*
	Infliximab	
	Adalimumab	
	Golimumab	
	Certolizumab pegol	
	Abatacept	*CD80, CD86*
	Anakinra	*IL1R1*
	Rituximab	*MS4A1*
	Tocilizumab	*IL6R*
DMARDs	Auranofin	*PRDX5, IKBKB*
	Azathioprine	*HPRT1*
	Cyclophosphamide	-
	Cyclosporine	*CAMLG, PPP3R2*
	Iguratimod (T-614)	*ELANE, PTGS2*
	Leflunomide	*DHODH*
	Methotrexate	*DHFR*
	Sulfasalazine	*ALOX5, PTGS1, PTGS2, PPARG*
	Tacrolimus	*FKBP1A*
	Temsirolimus	*MTOR*
	Tofacitinib (CP-690,550)	*JAK1, JAK2, JAK3*
Steroids	Prednisolone	*NR3C1*
	Methylprednisolone	*NR3C1*
	Desoxycorticosterone Pivalate	*NR3C2*
Others	Hydroxychloroquine	*TLR7, TLR9*

b



c



Extended Data Figure 7 | Overlap of all genes in the RA risk loci with drug target genes. **a**, Approved RA drugs and target genes. DMARDs, disease-modifying antirheumatic drugs. **b**, Overlap analysis stratified by immune-related and non-immune-related drug target genes. We made a list of 583 immune-related genes based on Gene Ontology (GO) pathways named 'immune-' or 'immuno-' and found that the majority of drug target genes (791/871 = 91%) were not immune-related. **c**, Overlap of all 377 genes included in 100 RA risk loci (outside of the MHC region) plus 3,776 genes in direct PPI with them and drug target genes. We found overlap of 19 genes from the 27 drug target genes of approved RA drugs (2.3-fold enrichment, $P < 1.0 \times 10^{-5}$). All 871 drug target genes (regardless of disease indication) overlap with 329 genes from the PPI network, which is 1.3-fold more enrichment than expected by chance alone ($P < 1.0 \times 10^{-5}$), but less than 1.7-fold enrichment compared with RA drugs ($P = 0.0059$). We note that this enrichment of drug–gene pairs was less apparent compared with that obtained from the expanded PPI network generated from 98 biological candidate genes (Fig. 3b).